DOCUMENT RESUME

ED 102 200 95 TM 004 257

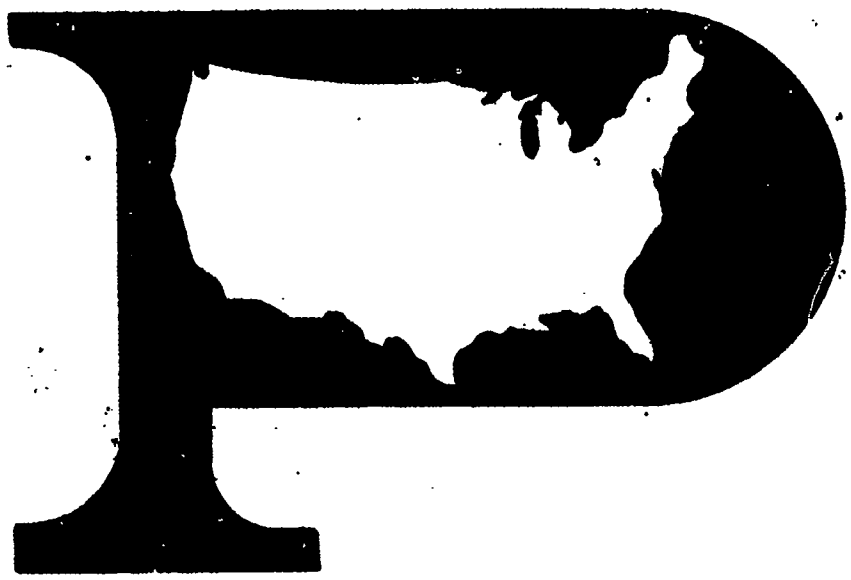
AUTHOR Andrews, Theodore E., Ed.
TITLE Assessment.
INSTITUTION New York State Education Dept., Albany. Div. of Teacher Education and Certification.; State Univ. of New York, Albany.
SPONS AGENCY National Center for Improvement of Educational Systems (DHEW/OE), Washington, D. C.; Office of Education (DHEW), Washington, D.C. Teacher Corps.
PUB DATE 74
NOTE 130p.
AVAILABLE FROM National Dissemination Center for Performance Education, Syracuse University, Syracuse, New York 13210

EDRS PRICE MF-$0.76 HC-$6.97 PLUS POSTAGE
DESCRIPTORS Certification; Costs; *Educational Assessment; Educational Research; Inservice Teacher Education; Lesson Plans; *Measurement Techniques; Models; Objectives; *Performance Based Teacher Education; Performance Criteria; Preservice Education; Program Evaluation; State Departments of Education; Student Evaluation; *Teacher Evaluation; Teaching Methods

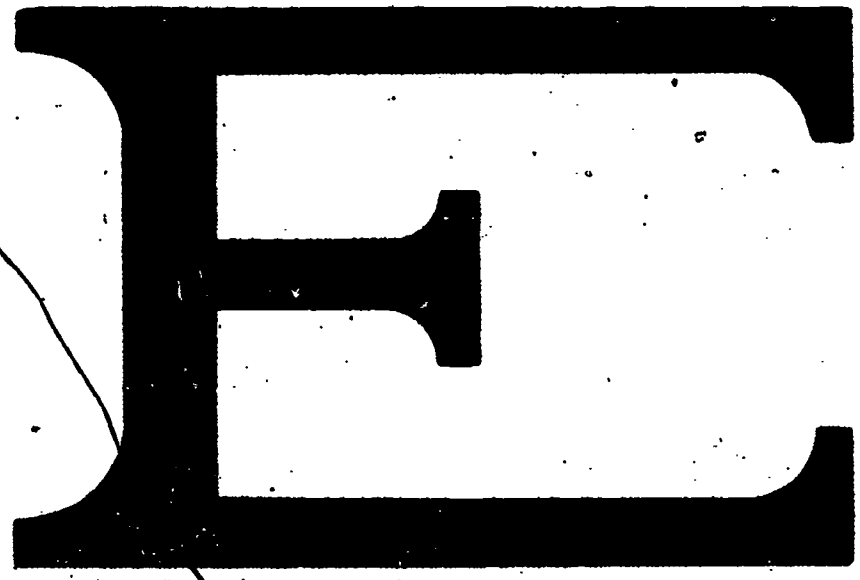
ABSTRACT

"The Role of the State in Performance-Based Teacher Education-Certification" by Robert Roth creates a context for viewing how state agencies are approaching performance education. Peter Airasian then explores the value questions that are at the heart of evaluation issues. Fred McDonald looks at "The State of the Art in Performance Assessment," and Barak Rosenshine lists recommendations concerning the research dilemmas. James Popham, in three papers, touches on the problems of selecting assessment systems, developing performance tests, and identifying minimal competencies. Del Schalock then details what occurs when "Moving from Conceptualization to Practice in Assessment." Concerning costs and teacher concerns, two educators, Bruce Joyce and Herbert Hite present their conclusions. Beatrice Ward discusses the cost factors involved in developmental work at the Far West Regional Laboratory. Finally two teachers, Sandra Feldman and Bernard McKenna, note their interest in the potential of performance-based teacher education and reveal their most serious concerns. (RC)

TM



PERFORMANCE EDUCATION



ASSESSMENT

THE UNIVERSITY OF THE STATE OF NEW YORK
THE STATE EDUCATION DEPARTMENT
DIVISION OF TEACHER EDUCATION AND CERTIFICATION
and
MULTI-STATE CONSORTIUM ON PERFORMANCE-BASED
TEACHER EDUCATION

ASSESSMENT

PETER AIRASIAN • SANDRA FELDMAN • JESSE H. GARRISON • HERBERT HITE
BRUCE R. JOYCE • BERT Y. KERSH • FREDERICK McDONALD • BERNARD McKENNA
W. JAMES POPHAM • BARAK ROSENSHINE • ROBERT A. ROTH
H. DEL SCHALOCK • BEATRICE A. WARD

Theodore E. Andrews, Editor

MULTI-STATE CONSORTIUM ON PERFORMANCE-BASED
TEACHER EDUCATION

CONTRIBUTORS

Peter W. Airasian
Boston College
Boston Massachusetts

Sandra Feldman
United Federation of Teachers
New York City, New York

Jesse H. Garrison
Oregon College of Education
Monmouth, Oregon

Herbert Hite
Western Washington State College
Bellingham, Washington

Bruce R. Joyce
Teachers College
Columbia University
New York City, New York

Bert Y. Kersh
Oregon College of Education
Monmouth, Oregon

Frederick J. McDonald
Educational Testing Service
Princeton, New Jersey

Bernard H. McKenna
National Education Association
Washington, D.C.

W. James Popham
University of California, Los Angeles
and
The Instructional Objectives Exchange

Robert A. Roth
Coordinator of Performance Education Project
New Jersey Department of Education

Barak Rosenshine
Bureau of Educational Research
University of Illinois at Urbana

H. Del Schalock
Teaching Research
A Division of the Oregon State System of Higher Education
Monmouth, Oregon

Beatrice A. Ward
Associate Director, Teacher Education
Far West Laboratory for Educational Research and Development

Foreword

The Multi-State Consortium is deeply indebted to the authors who helped shape this publication. Their willingness to share their experiences and opinions both with the consortium and with the readers of this volume is greatly appreciated. The consortium believes that the ideas contained herein, will not only interest readers but also influence policy decisions.

The consortium is particularly indebted to the United States Office of Education for its support. Created by a Title V grant, it has benefited both from additional fiscal resources and the direct involvement of representatives of the United States Office of Education. In particular, Teacher Corps (James Steffensen), National Center for the Improvement of Educational Systems (Allen Schmieder), and Title V (Stuart Dean) have been most supportive of its efforts and helpful in planning the meeting that led to the conference, the source for most of the papers appearing in this publication.

The Conference Committee included James Steffensen and Paul Collins (both of Teacher Corps), Bruce Joyce of Teachers College and the director of the consortium. The success of the conference and, it is hoped, of this publication, is in large measure a reflection of their efforts.

MULTI-STATE CONSORTIUM ON PERFORMANCE-BASED TEACHER EDUCATION

STATE	REPRESENTATIVE
Washington	Edwin L. Lyle
Oregon	James Wallace
Utah	Roger C. Mouritsen
Arizona	John Potts
Texas	Tom T. Walker
Minnesota	Patricia J. Goralski
Vermont	Robert Vail
New York	Vincent C. Gazzetta
Florida	Nell Kannwischer
New Jersey	Ward Sinclair

Administering State: New York

Project Director
Theodore E. Andrews
New York State Education Department

Stuart E. Dean
United States
Office of Education Representative

Syracuse University Coordinator
James Collins

Funded
by
United States Office of Education
Title V, Section 505, Grant

and

Syracuse University Projects
Supported
by
Teacher Corps
and
National Center For the Improvement of Educational Systems

Introduction

The Multi-State Consortium is pleased to present to interested readers the following papers.

We have titled this volume "Assessment" because of the importance we attach to the assessment problems. However, not all of the papers focus solely on assessment. In fact, it is difficult, probably impossible, to discuss assessment without quickly touching on such areas as research, costs, teacher evaluation, and program development issues.

Many of the articles were presented by the authors to the consortium at a series of meetings held in conjunction with the 1973 American Educational Research Association's annual meeting in New Orleans in February. Several of these have been significantly revised, and we have included other chapters that appeared appropriate in the context of this publication.

The authors' audience is primarily state education agency personnel and, in a larger sense, political decision makers. However, the logic and appeal of the selections is not limited to that group; anyone interested in the problems and potential of performance education should be engaged by these learned and personal opinions.

The first selection, "The Role of the State in Performance-Based Teacher Education-Certification" by Robert Roth, creates a context for viewing how various state education agencies are approaching performance education. Peter Airasian then explores the value questions that are at the heart of evaluation issues. Fred McDonald looks at the "State of the Art in Performance Assessment," and Barak Rosenshine makes a series of recommendations concerning the research dilemmas. James Popham in three separate papers, touches on the problems of selecting assessment systems, developing performance tests, and identifying minimal competencies. Del Schalock then details what occurs when "Moving From Conceptualization to Practice in Assessment."

The assessment emphasis now shifts to related topics: costs and teacher concerns. Little definitive information exists on the costs of developing and implementing PBTE programs and assessment systems. Two educators (Bruce Joyce and Herbert Hite) who have prepared cost analyses for their respective states (New York and Washington) present their conclusions. Beatrice Ward also discusses the cost factors involved in the developmental work at the Far West Regional Laboratory. Their collective conclusions are that programs will cost more, but that using what has already been produced will keep the costs within manageable limits.

Finally we offer two papers prepared by teacher representatives, Sandra Feldman and Bernard McKenna. Each notes their interest in the potential of PBTE while also revealing their most serious concerns. McKenna's paper concluded the AERA symposium and most fittingly concludes this volume as he recapitulates what has come before and reacts to what he has heard.

The Role of the State in Performance-Based Teacher Education-Certification

By

ROBERT A. ROTH

Certification and Education

The movement toward development of performance-based teacher education and certification programs has experienced rapid growth within the past few years. Currently there are approximately 30 states that are actively involved in the study of either performance-based teacher education or certification.[1] The purpose of this paper is not to discuss the pros and cons of a performance-based program, but to describe and examine the issues facing a state that has elected to move in this direction. Germane to the discussion are the changing roles and relationships that must be considered in order to plan effectively for implementation. Any scheme for development that has not considered and accounted for the essential underlying issues risks being nugatory in nature.

Initially it may be of value to recognize the traditional distinction between the process of teacher education and that of certification. Basically, teacher education serves a preparatory function, whereas certification selects those who are eligible for employment and provides them with a license. Certification traditionally has been a screening device, and it has been assumed that the state is the best agency to carry out this function. Discussion here will focus primarily on teacher certification as it has been a state responsibility.

It is interesting to note that there has been a great deal more resistance to performance-based certification than to performance-based teacher education. At a recent conference of the Regional Interstate Project held in Denver, the consensus appeared to support this distinction. Sandra Feldman, vice president of New York State Local AFT, stated "We do not oppose Performance-Based Teacher Education. The concept is a welcome one . . . *We oppose, however, a changeover to Performance-Based Certification at this time.*"[2] David Darland, with NEA, added,

> Most would agree to the importance of the performance dimension of educating teachers, but to establish one prototype of teacher education as the sole route to legal licensure is pure folly, especially in the absence of established evidence through viable research. To base advanced credentialling or renewing certifications on such a singular notion is ever more upsetting. This is not to decry experimentation with performance-based teacher education. Already some developmental approaches to performance-based teacher education appear promising, if not highly successful.[3]

The difference between certification and teacher education, however, varies significantly depending upon the particular certification model. Since there are many ways in which performance-based certification can be structured, criticisms should be centered around how the issues pertain to a given structure or definition of performance-based certification.

If one views certification (particularly the performance type) to be a testing procedure, then the distinction is clear and the meaning of the skepticism is more

[1] Robert A. Roth, "Performance-Based Teacher Certification: A Survey of the States," Trenton: New Jersey State Department of Education, Division of Field Services, December 1972.

[2] Sandra Feldman, "Performance-Based Certification: A Teacher Unionist's View," Performance-Based Education and Certification, report of the Regional Interstate Project Program, Denver, Colorado, July 18-20, 1972, Denver: Colorado State Department of Education, January 1973, p. 66.

[3] David Darland, "The Role of Professional Organizations in Performance-Based Teacher Education," Performance-Based Education and Certification, report of the Regional Interstate Project Program, Denver, Colorado, July 18-20, 1972, Denver: Colorado State Department of Education, January 1973, pp. 69, 70.

apparent. In this paradigm the preparing institutions are responsible for developing competencies in their teacher candidates and the state certifies a candidate's competence by testing him before issuing a license. Many difficulties associated with such a licensing procedure have been pointed out. One can argue, for example, that there exists no empirical base on which to construct a valid testing technique, particularly in view of varied teaching contexts. The problem is not the same with performance-based teacher education because there is a diversity of programs and flexibility to constantly develop and change the performance standards. Performance-based certification, it is argued, mandates only one way of teaching, seems more of a finality, and is less responsive to change.

In the approved program approach the distinction between certification and teacher education becomes less clear. When utilizing an approved program approach to certification in conjunction with performance-based criteria, as is frequently the case, the distinction may become even more nebulous. Supporting performance-based teacher education but opposing performance certification then becomes a tenuous position. Interestingly enough, the approved program approach is the predominant system in use. In 1971 it was reported that "at the present time, 36 states report extensive use of the approved program approach to certification, and in fact it has become the vehicle whereby forward-looking states have found the freedom to move in many promising new directions." [4]

It seems that the approved program approach to performance-based certification would be less susceptible to criticism than the state examination approach. In addition, it is the more common certification system currently in use and it provides a certain degree of freedom to explore new directions such as the performance-based model. The project, Improving State Leadership in Education, issued a report which concluded that "It would appear that the effective administration of a state-wide performance-based teacher certification system would depend almost entirely upon an effective system for program approval." [5] In view of these factors, the models for performance-based certification to follow will be within the context of the approved program approach.

Issues

In selecting a particular model, a number of important issues need to be considered. An essential question is what the role of the state should be in the certification process. There are at least two opposing viewpoints concerning the state's function. On the one hand, there are those who see the state as an administrative and regulatory body.

> The belief is that the state must improve its guardianship of the public interest by setting ever higher standards and developing more efficient systems of management. In one sense the state knows what is best."

This view of the state's role in certification is the predominant one currently in practice. It is a centralized approach with uniformity and standardization being the emphasis. Even an approved program approach could fit into this scheme if regulations concerning program content are specified. A performance-based certification system structured on the above tenets would specify teacher performance criteria for certification at the state level.

The opposing viewpoint on certification emphasized a decentralized system with more local control and a broader base for decision making and social change. In this strategy, "the state must promote change rather than mandate it and accept diversity as more responsive to the state's needs than mandated single standards." [7]

The competency approach could easily fit into this philosophy also by allowing teacher education programs or other professional agencies to develop their own particular sets of competencies. In fact, as Andrews points out in some places the competency movement "has been adopted as an attempt to reform the educational system by changing the locus of authority and thereby the way in which decisions are made." [8] One result of this is that a variety of stand-

[4] Improving State Leadership in Education, "Planning and Effecting Improvements in the Preparation and Certification of Educators," report of a Special Study, Denver, Colorado, April 1971, p. 7.

[5] Ibid., p. 15.

[6] Theodore E. Andrews, "Its Wisdom and Its Folly," paper presented to the Workshop on Problems of Competency-Based Teacher Education, Teacher Corps, State University of New York at Albany, May 1972, p. 7.

[7] Ibid.

[8] Ibid., p. 6.

ards appear, replacing the single set of state standards.

The implementation of a specific viewpoint of a state's role results in a number of ramifications inherent in the particular position. These consequences are, in effect, the underlying issues which impinge upon the decision to select a particular state role and therefore should be carefully considered.

In the centralized view of the state's role, a set of performance criteria would be established at the state level. These criteria may be developed by a state agency or through statewide improvement, the merits of which will be discussed at a later point. This standard set of statewide criteria can be utilized in an approved program approach or can be developed into a state testing instrument. Since the former has been determined to be possibly more advantageous, immediate discussion will follow in this context.

The approved program approach itself has been evaluated by some educators as being restrictive. Lierheimer has pointed out that,

> the colleges approved program must follow exactly the courses prescribed for state certification. Such a curricular requirement does not provide the freedom which colleges must have if they are also to be held responsible for the qualifications of the teachers they prepare.[9]

His remarks are made particularly pertinent to a competency-based program by substituting "performance criteria" for "courses" in his statement. Thus, lack of curricular freedom may result from a centralized state role with statewide performance criteria.

Curricular freedom extends beyond the right to decide on a particular set of courses. The freedom to experiment with innovative curricula also appears to be precluded by a rigid set of state performance criteria. The right of colleges to experiment becomes an important issue in the selection of a performance-based certification model.

The project, Improving State Leadership in Education, reported that critics of certification structures in general complain that "The rigidity of state requirements discourages flexibility and creativity in teacher preparation programs." [10] Further, "Ideally, the approved program approach would allow institutions to experiment and develop creative programs of teacher preparation and encourage innovation in teacher education within the framework of generally agreed upon goals." [11] An important part of this last statement is the word "generally." Generally agreed upon goals may still provide the freedom that Lierheimer is concerned about.

It would seem that the centralized view of the state's role with a standard set of specific performance criteria would be contrary to the intent of the approved program approach. Yet, performance-based certification appears to depend "almost entirely upon an effective system for program approval." [12] An approved program approach without highly specific criteria is an alternative.

Curricular freedom, the right to experiment, flexibility, innovation, and creativity in programs are issues related to the state's role that directly affect the teacher preparation institution. Other issues relate to the individual and the restrictions imposed by a specific set of performance criteria existing as state standards for certification.

McDonald relates that "The specifics of teaching competence will differ markedly depending on how we decide about the freedom each person will be given to choose the goals and means for his personal development and his life style." [13] At one extreme the teacher's services are sought requiring social skills, but at the other end he is an expert strategist requiring technical skills. A specific set of state standards may only permit one of these philosophies to prevail, as options may be impractical or even contradictory. Yet, one may argue that without state control contradictory standards could exist.

McDonald also raises a related issue. "Should we not consider whether a teacher has the freedom to define the nature of his service to students? Does he have the freedom to decide what will be required of him?" [14] Decisions on these questions clearly have implications for standardization of competencies and the role of the state.

An overriding concern with the performance criteria approach is that students will be boxed-in, forced to conform to a particular mold. It is argued by some that certification must provide for flexibility in person-

[9] Alvin P. Lierheimer, "Give Up the Ship," paper presented to the Massachusetts Advisory Council on Education, Boston University, February 3, 1968, p. 4.

[10] Improving State Leadership, p. 3.

[11] Ibid., p. 7.

[12] Ibid., p. 4.

[13] Frederick J. McDonald, "The Philosophical Problems of Competency-Based Teacher Education," Teacher Corps, State University of New York at Albany, May 12, 1972, p. 5.

[14] Ibid., p. 7.

ality, method, and philosophy (open classrooms, traditional, etc.). A specific set of standards at the state level does not provide for this flexibility. The decentralized state role does, as it allows diversity in programs and performance criteria.

The American Association of Colleges for Teacher Education in *Evaluative Criteria for Accrediting Teacher Education, A Source Book on Selected Issues,* asserts that "there are and should continue to be several philosophies of teacher education." [15] Will a centralized state role and specified performance criteria preclude varied philosophies of teacher education? Each state must examine its particular structure to determine whether or not this would occur.

Several other questions must be considered in relation to the development of a set of formance criteria at the state level. Can such criteria readily be changed? Can a standard set of competencies be developed to fit all teaching situations or must a number of sets of criteria be designed? In relation to the affective domain, Elam believes

> The competencies that are easier to describe and to evaluate are likely to dominate . . . The skills of teaching and the behaviors of a teacher which are difficult to learn and to evaluate often focus on the human aspects of teacher-pupil contacts.[16]

Can these performance criteria be established in the affective domain on a statewide basis, or are they situation specific and thus call for multiple standards developed at local levels? Will decentralization make the problem any easier to solve?

The arguments suggesting a need for an empirical base for performance-based certification but not teacher education were presented earlier. These arguments pertain to a certification system with a uniform set of standards at the state level, the centralized view of the state's role.

At a recent meeting of the American Federation of Teachers, the following statement was issued in a report.

> If state agencies begin to require the mastery of specific competencies as a prerequisite for certification, two dangers would exist. The first would be that pointed out earlier: non-validated knowledge and skill competencies as well as personal characteristics unrelated to true teaching effectiveness may be required, leading to certification standards perhaps even more non-relevant than those now existing. Second, pressure groups may be able to legislate requirements that attempt to define teachers and teacher behaviors into unacceptable patterns. A candidate could be required to fit the mold or not be certified.[17]

Perhaps general guidelines or a variety of standards developed by local groups or institutions would be less susceptible to these dangers. On the other hand, these groups may be just as likely to commit these errors.

In reference to establishing a minimum set of competencies at the state level, Andrews surmises that

> Evaluating the competencies demands a frame of reference, at its heart a set of values: I worry about states establishing value systems, thus the frame of reference must be diversified and most likely localized . . . Since we have a diverse population with varied philosophies, I believe a state should promote a certification system that expects diversity and challenges all to meet the highest level of accomplishment.[18]

Those who favor a uniform set of quality standards throughout the state, however, would seek the more centralized decision-making state role. Inequities among programs would thus be eliminated and employers would be assured that all certified personnel possess at least a minimum set of competencies.

In analyzing the models in terms of the issues, an important question should always remain in sight. In most cases it will not be a matter of whether or not a condition exists, but to what extent it exists. For example, to state that curricular freedom does or does not exist is merely an opinion that does not focus on the issue. The real issue is whether or not there is *sufficient* curricular freedom to satisfy those involved. Carrying the example to the other extreme, there may

---

[15] American Association of Colleges for Teacher Education, Evaluative Criteria for Accrediting Teacher Education: A Source Book on Selected Issues, Washington, D.C.: AACTE, 1967, p. 26.

[16] Stanley Elam. "Performance-Based Teacher Education: What is the State of the Art?" Performance-Based Teacher Education Series. No. 1. Washington, D.C.: AACTE, 1971, p. 19.

[17] American Federation of Teachers, AFT-QuEST Consortium Yearbook. Washington, D.C.: AFT, April 1972, p. 30.

[18] Andrews, "Its Wisdom," p. 12.

be circumstances that permit curricular freedom (or other conditions) to exist to such an extent that it destroys another essential or desirable element of a certification structure. The models must be scrutinized to determine if conditions are sufficiently provided for, but not overindulged.

Models

There are many ways in which a performance-based teacher certification system can be designed within the approved program approach. At one end of a continuum we have a very open system with maximum flexibility, whereas at the other end we have a highly structured and centralized approach (figure 1). There are, of course, many possibilities in between. Some of the models have been alluded to in the discussion of issues.

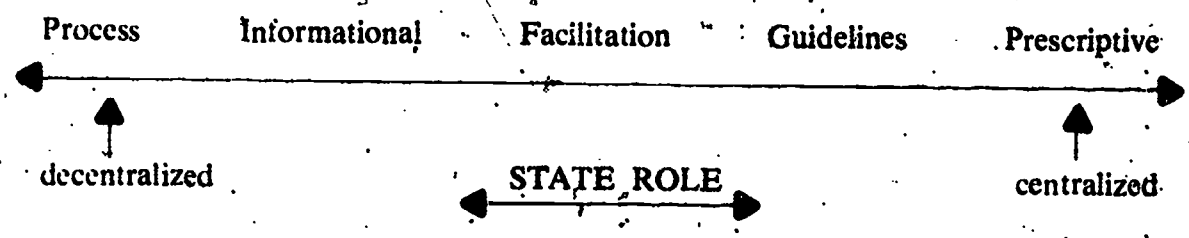


Fig. 1.
Continuum of Models
for
Performance-Based Teacher Certification
Approved Program Approach

The open-ended approach may be called the process model. In this system, the state does not determine the content of the teacher education program. Performance criteria are not established at the state level. The primary role of the state is to define the process for development of teacher education programs, stating who is to be involved and the nature of the involvement. In this model, the state plays a more decentralized role with more local control and a broader base for decision making.

Some states are now operating a competency-based certification system consistent with this model. The state of Washington is a primary example was the first state to adopt competency-based certification, and now has an operational program. A new set of standards for approval of teacher preparation programs became effective in Washington in September 1971.[19] Under these standards, preparation programs are to be developed and implemented by a consortium of agencies. Each agency designates its own representative(s) and clarifies with that (those) representative(s) his (their) authority in acting in behalf of the agency. The agencies in a consortium are colleges and universities, school organizations, and professional associations.

The professional association, determined by the total faculty of certified employees in a school organization in accordance with state law election procedures, has the responsibility of providing opportunity for input from all other specialized and subject matter associations. The school organization represents parents, local boards, and administration.

The consortium is charged with describing roles to be assumed by the person to be granted a specific certificate and with identifying and stating the rationale for the competencies required of persons who plan to perform the described roles. The certificates will be issued by the state through an approved consortium program. These standards are themselves process and performance standards.

In reference to this model, it would be of little meaning to support performance-based teacher education but not performance-based certification. One merely provides for the other; hence, they become part of the same process. As noted earlier, the necessary task is to examine the various certification models in terms of the issues rather than compare certification with teacher education.

Clearly, this state has moved toward a decentralized structure with more local control, a broader base for decision making, and diversity of standards. Performance standards are more readily changed with feedback, and probably less resistance would be encountered in the state. This model values optimum freedom for the preparing institution in relation to curricular decisions, flexibility, and creativity. In terms of the in-

[19] State of Washington, "Guidelines and Standards for the Development and Approval of Programs of Preparation Leading to the Certification of School Professional Personnel," Olympia, Washington: Superintendent of Public Instruction, adopted July 9, 1971.

dividual, there is the possibility, depending on the program, for freedom to define goals, for flexibility in personality, method, and philosophy. Reflecting this viewpoint, William Drummond, a former associate in the Washington State Department of Education, urged that "State departments of education, therefore, should foster creativity and intellectual freedom and promote programs of teacher education which support and cherish uniqueness and individualism." [20]

The Washington model, therefore, also rejects the regulatory role of a state department of education. Wendell Allen, as Washington's assistant superintendent of public instruction, concluded

> To emphasize this regulatory role is to protect the status quo. When the rule is the thing, change must come before there can be a new rule. There is danger in this circumstance that the major energies of the agency will be spent on administrative rather than leadership functions.[21]

An essential point to note is the prevalence of multiple standards, lack of uniformity, less legal need for an empirical base, and no single set of standards. Should all of the above factors be deemed advisable, then a particular state might select this model.

New York has envisioned a very similar type of program.[22] Four process standards have been established to be utilized for the development of pilot projects. The standards require the establishment of a policy board made up of representatives of teachers, school districts, colleges, and teacher education students. This group considers the objectives of the schools involved, the competencies teachers need to be successful in that environment, as well as those qualities desirable for all teachers, and acceptable evidence for attainment of competencies. The policy board then will establish individualized programs for the preparation of teachers to meet these criteria. Finally, a management system must be established. Trial projects may be designed for initial or continuing certification or both. The State Department of Education will exercise its legal responsibility for program approval. Note the decentralized role and the belief that performance criteria are mostly situation specific.

Vermont [23] has expanded the decision-making base to local school districts. A local school district may develop a program for the inservice training and professional advancement of its staff and may apply to the State Department of Education for approval to recommend issuance and renewal of all certificates at the local level. The appropriate certificate will be issued by the State Department of Education.

The local district must submit evidence that the teachers, school board, and administrative personnel have participated in the planning and development of the program. The local program must include provision for job description, task analysis, and performance criteria for all education personnel. An approved program approach is in effect for college teacher preparation programs.

Washington, New York, and Vermont are case studies that fall into the process model. Local decision making characterizes these attempts, assuming what is acceptable in one situation may be unacceptable in another.

Moving slightly along the continuum away from the process model but within the local decision-making framework, there is a model suggested by Lierheimer [24] which we may call the informational model. The central thesis of Lierheimer's proposal is that the state's role is not to make judgments but to maintain records. He suggests the students be tested over a multitude of factors including actual teaching performance. There is a possibility here for utilization of performance criteria, but the testing is not done by the state.

Decentralization is emphasized in this approach with local school teams conducting the evaluation of the competence of potential teachers. Ultimately, the agency to decide on teacher performance for licensing purposes would be the school. The function of the state is to monitor the local evaluation but not impose state standards. Although evaluation systems would be approved by the state there would be no uniform techniques for verification of classroom performance. The state office would maintain a data bank on all teaching personnel in the state.

[20] William H. Drummond. "Conference Commentary" in The Seattle Conference: The Role of the State Department of Education in Teacher Education, Olympia, Washington: State Superintendent of Public Instruction, 1967, p. 74.

[21] Wendell C. Allen, "State Government and Teacher Education —A Different Role for the State Education Agency," in The Seattle Conference. p. 78.

[22] New York State Education Department, "A New Style of Certification." Albany, New York: New York State Department of Education, Division of Teacher Education, March 15, 1971.

[23] Vermont State Department of Education, "Regulations Governing the Certification of Educational Personnel," Montpelier: Vermont State Board of Education. Department of Education. July 1. 1971.

[24] Lierheimer, "Give Up The Ship."

A unique feature of this model is that the state accumulates information on an individual but makes no decision in reference to competence. The major role of the state is to provide resources. The local district is provided with the information, and it is at this level that decisions are made as to whether the individual's competence fits the particular situation. The underlying assumption is that values and competencies are situation specific and hence require local evaluation. Currently, there are no states utilizing this informational model. Again, analysis of the model should be made in terms of all the issues identified earlier.

This model can be modified to interject more state control and greater uniformity. Minimum standards could be set by the state for the various competencies or groups of competencies. These minimum standards would be established for certification purposes. The state would still maintain its individual data bank and local districts could use the information for hiring purposes. This modified model would be farther along the continuum in terms of state control and decision making.

Another open-type model which does not provide quite as broad a decision-making base is being developed by the state of Florida. In this case consortia are not designed for purposes of initial certification although inservice programs are developed by local districts. This "facilitation model" utilizes the college approved program approach commonly in practice among the states.

The program approval regulations are somewhat process in nature indicating prescribed activities, but they are content standards as well, identifying courses necessary for certification. There are alternatives to the content regulations which provide for performance-based programs.

> An institution may, instead, specify the competencies which its graduates will be expected to demonstrate, identify the procedures by which those competencies will be measured, and then develop a program which leads to those competencies. Once such a program is approved, its graduates will receive regular teaching certificates with no penalties. Institutions are now being encouraged to develop competency-based programs.[25]

In this model control is in the colleges, but direction is provided by the state. The colleges develop their own competencies, but these are consistent with state course requirements. There is additional direction and stimulus provided by the state, however, which facilitates development of such programs. The state is compiling a catalog of teaching competencies which will eventually be validated through research. These competencies, or performance criteria, will be provided to the colleges to facilitate their program development. These particular criteria, however, will not necessarily be mandated and certainly all will not be required of a given institution. Other facilitating procedures by the State are assembling of training materials based on performance criteria and staff development for teacher trainers. The emphasis is on facilitation. Decision making is somewhat diffused but the role of the state is stronger than in previous models. The facilitation model presents different responses to the issues.

The remaining two models to be discussed can be grouped under a heading of central decision making. The first two models, you may recall, were local decision-making types, with the facilitation model being somewhere between. These last two models are at the other extreme end of the continuum.

One approach to performance-based certification is to establish performance criteria at the state level. This approach supports a strong state role and a uniform set of standards. It guarantees that each certificated individual has at least a minimum set of competencies. These criteria could be utilized as a state test or part of an approved program. The focus here, however, is on the approved program approach.

The manner in which these criteria are stated significantly affects the impact they will have on teacher education programs and the role of the state. The performance criteria can be stated in generic terms which then serve as guidelines for further specification by teacher preparation institutions. This guidelines model increases centralized authority yet does provide a certain degree of participation on the part of the colleges or consortia.

As an example of competencies consistent with the guidelines model, it might be required that the teacher candidate demonstrate the ability to diagnose areas of student deficiency, maintain a classroom environment which motivates students to learn, plan an instructional unit, employ a variety of instructional techniques, etc.

Utah[26] recently adopted at the state level a set of

[25] Florida Department of Education, "The Florida Program for Improving the Training, Evaluation, and Licensure of Educational Personnel," draft number 2, Tallahassee: Educational Research and Development Section, April 7, 1971.

[26] Utah State Board of Education," Recommended Proficiency Guidelines for Media Endorsements," Salt Lake City: Utah State Board of Education, 1972.

Robert A. Roth

performance criteria for instructional media, some of which approximate the guidelines model type of criteria. Prerequisite to a Basic Media Endorsement are a bachelor's degree and a teaching certificate. An examination for proficiency conducted by a recommending institution (with an approved certification program) is then administered. The recommending institution is free to determine how the competency will be demonstrated or ascertained, but a candidate may request an opportunity to demonstrate a competency whenever he feels he is ready. Competencies may be demonstrated one at a time. Candidates who perform satisfactorily will be considered as having met the endorsement requirement regardless of the route taken to obtain the competency.

Proficiency must be demonstrated in five areas. Some examples of performance standards are as follows:

—Using media selection tools of his choice, the candidate will identify the tools he has selected and include a rationale for the choice of each.

—The candidate will explain what one would do to select new subject headings for materials which are not considered in "Sears List of Subject Headings."

—The candidate will demonstrate proficiency in mounting pictures by producing one acceptable example of the following:

(1) dry mount on a hard surface, using dry mounting tissue

(2) dry mount, using dry mounting cloth

(3) rubber cement mount

(4) laminate with thermo copy machine, adhesive acetate, or heat press.

At the extreme end of the continuum we have what can be termed the "prescriptive model." In this system the state provides very specific performance criteria (behavioral objectives) which are utilized by the colleges as objectives and evaluative criteria. This is the most dominant of the state roles within an approved program approach with an emphasis on the administrative and regulatory function of a state education agency. Uniformity in certification with a single set of standards is the essential feature.

The state of New Jersey is currently studying the feasibility of such a performance-based certification system. Specific performance criteria are being developed for use on a statewide basis as certification standards. How specific these will be has not as yet been determined. There are two unique aspects to the New Jersey approach, however, that broaden the base of decision making. The performance criteria are being developed by task forces composed of a crosssection of educators from across the state, representing teachers, administrators, college students, college professors, the State Department, and the various professional associations. These criteria, therefore, are not developed by the State Department but represent a consensus of professional educators in the state. In addition, evaluation of prospective teachers may involve schools, colleges, and professional associations, a resemblance to the consortium idea.

Clearly, the statewide involvement in development of criteria adds considerable power to the approach. It presents a decided advantage over development of criteria by a state department or even a college or university. It appears to have greater validity and is more likely to find statewide acceptance. Significantly, "it has been generally agreed that whoever determines certification requirements controls the program of preparation." [27] Thus, in this instance, control is more in the hands of the total profession.

The guidelines model can be developed by the same method. The difference between the two models then lies in the specificity of the criteria. How does this difference relate to the issues, and how do these two models compare with the open end of the continuum in terms of the issues?

[27] Improving State Leadership, p. 5.

In the process model, teacher preparation institutions have maximum curricular freedom. The guidelines model allows the institutions the opportunity to develop the specific performance criteria while the prescriptive model does not provide for this. A comparison of performance criteria with traditional course list standards may be of value at this point. A course in tests and measurement is a familiar requirement in the course list system. The guidelines model would require competencies that are somewhat more specific, such as ability to evaluate student performance and ability to develop tests. The prescriptive model, however, would list a number of specific performances such as ability to formulate essay (multiple choice, etc.) test items and analyze tests for validity. Also, the *evidence accepted* that the performance had been achieved would be provided. Continuing our comparison, if used in a course list system, a prescriptive model would list the things that should be taught in a tests and measurement course rather than leaving this to the college.

Andrews has stated that "a required set of performance criteria could be just as moribund as rigid course requirements have been in the past." [28] It appears that the more specific the criteria the less freedom that exists. Recall that the approved program approach works "within the framework of generally agreed upon goals." [29] The possibilities for creativity through innovative programs can be achieved in the design of means to achieve the objectives, but not through alternative objectives. Two basic questions are at hand. First, is curricular freedom seen as being of value; and second, does a prescribed set of specific performance criteria significantly limit this freedom? A related question is whether or not the guidelines model offers a great deal more freedom than the prescriptive model.

A concern similar to the question of freedom is diversity. The process model allows, and even encourages, diversity among programs. Those in favor of diversity argue that there are varied philosophies of education requiring different teaching models. Any set of performance criteria is based on a theory of teaching and the teaching-learning process. Although not always articulated the purposes of teaching are inherent in the criteria.

In the process model several teaching philosophies exist simultaneously with validation and development being ongoing processes. A set of specific criteria, however, relies on one teaching model and also establishes a particular value system. The problem here is that there is no empirical base to lead us to the correct model. As noted earlier, lack of an empirical base is a primary concern with performance-based certification. With a variety of program types, it can be argued, we recognize the developmeatal state of our knowledge base, whereas a single model seems a finality and demands empirical validation before being adopted. This accounts for the support of performance-based teacher education instead of certification.

Another point made by those favoring diversity is that performance criteria are situation specific. There are numerous contexts for teaching, both in terms of environment and educational philosophy. This requires different sets of competencies, at least in terms of the general situations (not for every school, etc.). There may not be enough in common to establish at least a minimum core of competencies at the state level. Washington, Vermont, and New York appear to believe in this as evidenced by their process models.

All of the above factors suggest multiple standards and diversity of programs. The initial question is whether these are valid concerns. The other position argues for more standardization and quality assurance. Inequities among programs are diminished. Certainly, the prescriptive model adheres to the latter viewpoint. The guidelines model does provide a certain degree of variability in that each institution can define the specific criteria to fit its needs. The prescriptive model insists on a single standard; the guidelines model offers some degree of multiple standards although minute when compared to the process model.

A frequent criticism of competency-based programs is the problem of writing performance criteria in the affective domain. This problem becomes amplified as we move across the continuum toward the prescriptive model. As an example, the guidelines model might require competence in developing teacher-student rapport. Each teacher preparation institution would be provided the freedom to determine not only how this might be developed but how it might be judged to exist. The prescriptive model, however, would specify the performance criteria necessary to achieve this, such as "uses student names," or "smiles or acknowledges student responses by nodding." The question is whether or not such criteria can be written on a statewide level. Ignoring the affective domain and concentrating on the cognitve and psychomotor would not be a viable alternative.

The reader may recall the issues raised concerning

[28] Theodore Andrews, New Directions in Certification, Washington, D.C.: Association of Teacher Educators, 1971, p. 10.

[29] Improving State Leadership, p. 7.

the rights of the individual as suggested by McDonald.[30] Are there opportunities for flexibility in personality, method, and philosophy? What about the right of the individual to define his own goals? Rackley and Miller, as members of the Pennsylvania State Department of Education, stated that

> Individual differences are not taken into account in blanket certification standards. We are convinced that the improvement of teacher preparation must take place at the point of initial preparation . . . with attention directed to individual needs within the context of general certification requirements.[31]

The process model provides for individual flexibility, and, there are functioning programs which operate on these premises. The prescriptive model precludes much of this, at least in terms of the specific criteria required by the state. The individual does not have the freedom to define his own goals, but he may have the opportunity to select his own method of achieving the objectives. Again, those favoring a uniform set of standards would find individual selection of goals to be undesirable and detrimental to certification.

The guidelines model may provide a certain degree of individual choice but within the boundaries defined at the state level. The general objective must be accomplished, but the specifics can vary with the individual. The manner in which one wishes to develop teacher-student rapport, or plan for a lesson can vary significantly from another individual's method. The basic question is not just one of uniform standards versus flexibility but the degree of each that is desirable.

Alternatives to Approved Programs

The discussion of issues and alternatives has thus far been limited to the approved program approach to state certification. Approved programs referred to those developed by colleges alone or by consortia. The evidence presented earlier in this paper suggest that approving programs is the more viable approach to performance-based teacher certification, and some specific criticisms of the state testing approach were described.

There are some teaching areas, however, that find themselves less rigidly tied to college preparation programs and thus are more amenable to alternative approaches. The area of vocational education, for example, is somewhat unique in that it usually relies on experienced professionals in the various trades to enter the teaching profession. There are other areas, such as music, that also require specific skills unique to the particular profession. Educational fields such as these warrant consideration of alternative approaches that are not necessarily bound by college degree programs. These different approaches are not necessarily limited in application to the special teaching areas mentioned, however, as the alternatives may be utilized for any teaching field if desired.

A commonly discussed alternative to teacher certification is the establishment of a state testing procedure. There are several ways in which this can be implemented, some of which will be described here. A cogent argument against this approach (which was pointed out earlier) is that there exists no empirical base on which to construct a valid testing technique, particularly in view of varied teaching contexts. The predictive validity of any such examination device would have to be established.

It is again important to consider how the state testing models reflect the various issues. Questions about curricular freedom, individual freedom, and varied teaching philosophies should not be forgotten. The state testing approach to certification offers radically different responses to the issues when compared to the models within the approved program approach.

The informational model suggested by Lierheimer can easily be modified to fit a state testing procedure. A set of behaviorally stated competencies could be formulated as certification descriptors. A teacher candidate's degree of accomplishment of each of the criteria could be indicated to form his competence profile. Minimum standards established for certification could be set by the state for each criterion or group of criteria. A system could be established (total score, weighted scores, etc.) to determine the individual's eligibility for certification. The state would still maintain its individual data bank and local districts could use the information for hiring purposes.

An important modification of the Lierheimer informational model is that not only are minimum levels established for certification, but the testing of the candidate to determine his achievement of each criterion is done by the state, not through an approved program approach. The control of standards and verification of accomplishment reside in the hands of the state.

The modified informational model is but one variation of the state testing concept. Any outside agency or group of evaluators could be designated by the state to

[30] McDonald, "Philosophical Problems."

[31] J. R. Rackley and Norman Miller, "Broad Policy Concerns and Direction for a State Department of Eduation in Teacher Education," in The Seattle Conference, p. 15.

carry out the testing function. There is an opportunity to involve members of the profession in both development of criteria and service on evaluating boards or teams who certify individuals. Instead of a profile, verification of minimum competence might be all that is necessary. Differentiated certification could be based on different degrees of accomplishment or even different types of criteria. Evaluating boards or teams could again be used throughout the entire process.

It is generally assumed that the evaluation for certification would be done in a live classroom situation. An alternative would be to establish testing centers where specific skills would be evaluated such as those found in microteaching. This might be particularly useful for initial certification due to the inequities in student teaching situations. Students could also be used in test centers similar to the laboratory schools. This would provide a more controlled situation and fewer variables would enter into the evaluation.

A combination of evaluation in student teaching settings and controlled laboratory situations is also an alternative. This might be built into a system where a recommendation from a preparing institution (college or consortium) in addition to testing in a center would be necessary parts of the process for certification. The variations to this testing approach are too numerous to be included in this discussion.

Epilogue

Each model must be considered carefully in terms of the issues identified. Certainly, there are other issues to be accounted for which were not discussed here. The idea of certification levels was not presented in this paper and could by itself be an entire area of discussion with direct bearing on the selection of models. Another important question is whether or not to use student outcomes as an indication of teaching competence. Concerns of a practical nature such as cost, overall feasibility in terms of management, state size, diversity, and available resources are examples of other issues. The questions raised here were more of a philosophical nature and are pertinent to decision making.

The models described were identified as being along a continuum. This implies that there are many other models which can be considered, but they most likely will differ from these models in degree rather than basic type. Perhaps a system can be developed with positive elements from several of the models described here. It may also be possible that more than one model can be in operation at a given time, particularly if one accepts the notion that certain areas require or more readily fit into a state testing approach while all other areas fit one of the approved program models. The overriding concern is which model or models best serve the purposes of certification.

Performance-Based Teacher Education: Evaluation Issues *

By

PETER W. AIRASIAN

Introduction

In this paper I shall consider two levels of evaluation in performance-based teacher education models. The first concerns the evaluation of the basic concept of performance-based teacher education. The second relates to evaluation issues within the context of an ongoing performance-based approach. I shall argue that the important evaluative issues are *not* those of selecting appropriate and objective measuring techniques but rather involve the types of judgments which dictate who will be evaluated and on the basis of what performances. Those who are seeking cookbook, practical, "how to" answers from my presentation will be disappointed. My purpose is to argue that "how to" problems are predicated upon a prior set of evaluative decisions, decisions ultimately more important and powerful than any set of evidence gathering techniques, no matter how objective or complete these are.

The majority of evaluation issues raised by performance-based teacher education are not new. What is new and significant about performance-based teacher education is that for the first time, systematic evaluation of teacher training models and products is called for. It is the required synthesis of evaluation problems which heretofore have been treated largely individually which makes evaluation an issue in performance-based teacher education.

Evaluation

We must start with a definition of evaluation, since the term has multiple meanings according to the context within which it is considered. Perhaps the most commonly understood meaning of the term is "a judgment of the extent to which learners have mastered the objectives of instruction." While this definition is adequate within the context of many instructional programs, it is limited in three senses. First, it narrows attention to the intended outcomes of instruction. Second, it largely ignores information gathering about the process of instruction. Finally, it makes the learner the sole object of evaluation. To limit evaluation of performance-based programs to student evaluation is to overlook many of the nonlearner aims and assumptions inherent in the approach. It is important, particularly in the early stages of acceptance and implementation, to articulate all the evaluative issues inherent in performance-based programs. To undertake this task, evaluation must be recognized as encompassing a more general purpose than simply determining learner mastery of course outcomes. In the context of this paper, then, evaluation will refer to "a value judgment of merit or worth." Clearly this is a more general definition than that discussed above. It is meant to be. Omitting referents such as the learner, the curriculum materials, or the instruction from the definition focuses attention upon the process involved in performing an evaluation; that is, making a value judgment. It also suggests that whenever judgments are made, regardless of the referent, evaluation has taken place.

The heart of the evaluation process, then, is valuing. Data gathering, be it "hard," objective data or "soft," impressionistic data, is not evaluation. Evaluation takes place when data are compared to some standard or norm and a decision about pass or fail, accept or reject, or good or bad is made. For example, the fact that about 85 percent of the age cohort in the United States completes secondary school does not, in and of itself, convey value. Some of my colleagues, when presented with this fact say "Isn't it good that more young people in America graduate from secondary school than in any other country in the world?" Other colleagues respond "*Only* 85 percent—and I thought we were doing much better." Or, to select an instance

* Invited paper, Multi-State Consortium on Performance-Based Teacher Education, New Orleans, Louisiana, February 26, 1973.

closer to performance-based teacher education, consider a prospective teacher who teaches 60 percent of a group of sixth graders the difference between simple and compound sentences. Is mastery by 60 percent of the class grounds for congratulating or chastising the teacher? It is only when data are judged in terms of standards or norms that one can ascribe value to performance. There is, in essence, a difference between measurement (gathering data) and evaluation (placing a value on the performance).

Despite our wishes to the contrary, decisions about what is good, valuable, worthy, and desirable are made not in the research domain, but rather spring from our individual philosophies and frames of reference. If I were to ask you to cite three examples of school practices which are based firmly in established research findings, I dare say that you would be hard put to respond. Arguments for and against performance-based education reside, and will continue to reside, in the value domain. The questions asked are not whether one can state performances in behavioral, measurable terms but whether one should state them in any terms; not whether teachers can be evaluated on the basis of their students' performance, but whether they ought to be. These are issues which are based in philosophy and value orientations, and it is in these frameworks that evaluation centers. As a consequence, it is important to consider evaluation issues as they relate to arguments for and against the basic idea of performance-based teacher education.

Evaluation of the Performance-Based Model

The antecedents of performance-based teacher education are many, but they all revolve around a single central theme: accountability. There is a serious breakdown in the interest and ability of American citizens to support education. Educators are being forced to accept responsibility for their activities and products. Evidence of success—and if not success at least efficiency —is becoming a prerequisite for continuing financial support. The "cult of efficiency and rationality" is with us again. "Demonstrated value for dollars expended" and "It is time to make education scientific" are the watchwords.

Now, my experience tells me that these are the watchwords of legislators and administrators, not of classroom teachers. Accountability is, of course, always threatening to those who are to be held accountable. There is growing evidence (e.g., Jackson, 1968; Good and Brophy, 1973) that classroom teachers simply don't think of accountability in the same rational, efficiency-oriented terms as their administrative superiors. Moreover, classroom teachers perceive accountability, rightly or wrongly, primarily as a method of pointing the finger of blame at individuals, rather than at a myriad of interrelated and difficult-to-define factors.

Performance-based teacher education is here today not as the result of a ground-swell of support from teachers, but because administrators and politicians believe it is a more rational, efficient, and accountable method of training and certifying teachers, as well it may be. Performance-based programs are with us because they are believed to be better by those who have some say in the matter, not because research evidence overwhelmingly indicates their superiority. The decision has proceeded from the top downwards. In short, the values implicit in a performance-based approach fit closely with the values of legislators and administrators. An implicit evaluation has, therefore, already taken place. The evaluation was not based upon hard data, but rather upon a comparison of the perceived values inherent in performance-based versus other approaches to teacher education. In a number of states it is clear that the performance-based approach has been evaluated as superior.

The benefits claimed for a performance-based approach clear and public specification of ends, emphasis on exit not entry behaviors, individualized instruction, accountability, etc. are well articulated. It is the expectation of these outcomes which affords the justification for current programs. Or, perhaps more appropriately, it is the value placed upon these outcomes which affords the justification. However, in accepting the performance-based approach, one is making a series of assumptions about teaching, teacher training, and teacher evaluation. It is important to identify these assumptions because when performance-based teacher education is criticized and discussed, it will be primarily on the basis of its inherent assumptions

Before elaborating these assumptions, it is appropriate to pause here to reemphasize the main point of my argument thus far. In the social sciences, arguments for and against innovations and practice are based primarily upon the perceived values embodied in the innovations and practices. Because the social sciences have few, if any, paradigms or models which are universally accepted, social scientists wage their battles at the level of first causes or starting points. The physical sciences are different from the social sciences insofar as the physical sciences contain laws, theories, and principles to which all physical scientists subscribe. One phy-

sicist may not agree with another's application or research in the area of relativity theory, but both accept the basic axioms of the theory. Such is not the case in the social sciences. We are not at the stage of arguing about applications, but rather about starting points. We have five theories of personality, not one; three theories of what the "good" teacher is like, not one; four learning theories, not one. Each of us accepts different starting points about the nature of education, of man, or of the good. However, despite our rush to emulate the physical sciences, one of the most powerful aspects of the social sciences may be its pluralism. When we argue and criticize it is more often about assumptions than about application or practice. To a humanistic educator performance-based teacher education is anathema regardless of whether it succeeds or fails because it emphasizes the wrong things. The point of this rather lengthy digression is that evaluation of performance-based teacher education by those outside the fold will center upon the assumptions inherent in the performance-based approach as viewed from their value premises. Adherents to the movement require no evaluation of the central assumptions. They have already evaluated and endorsed them as good. Note that while performance-based proponents acknowledge the need for research about teaching activities and outcomes to improve the approach, few argue for a test of the basic idea itself. The basic idea has already been evaluated and accepted.

What, then, are the assumptions inherent in the performance-based approach? Major assumptions number five and are as follows:

1. There exists a set of performances which is important for all teachers to possess.
2. It is reasonable to identify, define, and set standards for the relevant performances.
3. Once identified, we possess the knowledge and skill to teach the relevant performances.
4. Teaching can be characterized as the sum of the defined performances.
5. The performances are measurably related to student learning.

The fact that I have called these five propositions assumptions does not imply that proponents of performance-based teacher education consider these to be assumptions. In fact, the extent to which any state, locale, or university has adopted a performance-based approach is the extent to which these propositions are not assumptions, but givens. While thoughtful advocates of performance based teacher education recognize the need for more research in areas touching on these five areas, most advocates accept the statements as facts, awaiting certification by research. To critics of the performance-based notion, however, the five statements are regarded as assumptions and will be the focal points of criticism and attack.

To illustrate my point, consider an article by Arthur Combs, director for humanistic education at the University of Florida, which appeared in a recent issue of "The Journal of Teacher Education" (1972). The title of the article is "Some Basic Concepts for Teacher Education." I have selected this article as an example not because I favor the arguments it advances, but rather because it happened to be sticking out kitty-corner from a heap of similar articles on my desk. Strictly random access to the heap would have served the same purpose. Here are two assertions advanced.

1. "The production of an effective teacher is a highly personal matter, dependent primarily upon the development of an appropriate system of beliefs." (p. 286)
2. "Effective teacher education must concentrate its efforts upon meanings rather than behaviors." (p. 287)

Combs goes on to argue, "It is conceivable that requiring a teacher education program to define precisely the behaviors it hopes to produce may be the surest way to destroy the effectiveness of its products by concentrating everyone's attention on the wrong dynamics" (p. 288). It is clear that Combs' value orientation places him in a teacher education camp far different from that of performance-based adherents. If a performance-based approach could be shown to achieve its aims, I would expect Combs' reaction to be, "Fine, but that's not what teaching is all about."

Returning to the five assumptions stated above, it is possible to consider each individually and to raise questions about its appropriateness. One could question the extent to which teaching is actually a science which can be analyzed and described in precise terms. Given current knowledge about teachers and teaching, one could raise questions about the ethical issues involved in identifying relevant teacher competencies and levels of performance. Even if certain performances were known to be related to teaching success, the question of whether such performances can be taught to prospective teachers is not at all clear. Within the context of performance-based teacher education, such issues are tagged as "problems to be solved," not as "considerations to be weighed in determining whether we should have performance-based programs."

Performance-based teacher education advocates have

proceeded from a belief system which accepts these five assumptions as givens. As a result, what is perhaps the most important evaluative question about performance-based teacher education—whether or not it is better than current practices and should be implemented—has been answered in many states. The likely impact of this value judgment on states, teacher education institutions, prospective teachers, and students outweighs the impact of evaluations carried out within any ongoing performance-based program. Once one's belief system accepts the need to go ahead with performance-based teacher education and certification, problems reduce to those concerned with implementation and "spreading the gospel." It was possible to have approached the issue of performance-based teacher education as a societal experiment, emphasizing planned intervention and evaluation of the basic notion itself (Campbell, 1969). However, it should be noted that this experimental approach also proceeds from a belief system, one which accepts a rational, planned, gradual approach to societal change.

Regardless of the basis on which the decision to institute a performance-based approach is made, however, once it is agreed to commence, issues related to evaluation within ongoing programs arise. It is to these problems I shall now direct attention.

Evaluation in Performance-Based Teacher Education Programs

A multitude of models which integrate both student and curriculum evaluation concerns of instructional programs have been advanced in recent years (e.g., Airasian, Madaus, and Rakow, 1972; Provus, 1969; Stufflebeam, 1971). These models describe various types of evaluation, the points in instruction where the types are most relevant, and the kinds of judgments which can be made on the basis of the evaluations. Most models are specifically designed to help individualize instruction within the context of programs which emphasize predefined objectives or competencies. Because such models exist, I shall not attempt to suggest a full-blown evaluation model for performance-based programs. Rather, I shall direct may attention to a few evaluative issues which I feel are of major importance for any performance-based teacher education model.

Defining Teacher Competencies

While some observers argue that the most powerful individuals in a performance-based approach are those who ultimately certify performance or competency, I would argue that the most powerful individuals are those who frame the competencies to be attained. These are the individuals who explicitly define what is a good teacher. The decisions of these individuals color the selection of learning experiences as well as the evaluative techniques and criteria. In the performance-based approach, which proceeds from identification of ends to selection of means to obtain these ends, it is the ends which are paramount. The rationale for a program, its learning experiences, standards, and certification practices rest upon the performances defined as needed by the good teacher. Parenthetically, there is a vast literature on the pros (e.g., Bloom, Hastings, and Madaus, 1971; Block, 1971; Popham, 1968; Tyler, 1934, 1950) and cons (e.g., Atkin, 1968; Doll, 1971; Eisner, 1966) of an approach to education which proceeds from clear articulation of ends to selection of means. Seldom does this literature appear in the context of discussions about performance-based teacher education.

It is at this first stage involving the definition of competencies that evaluation should play a major role. Just as controversy about performance-based teacher education generally will center upon its assumptions, so will controversy about any given performance-based program center on its specific definition of the good teacher. In short, one of the primary domains of evaluation within a given program will be its goals; i.e., the competencies it seeks to teach. That this will inevitably be the case is a function of the fact that not everyone defines a good teacher in the same way.

Wise administrators of performance-based programs will build evaluation of their program's defined competencies into their planning efforts. One useful model for such evaluation is suggested by the National Assessment of Educational Progress study staff. The NAEP sought to provide nationwide census-like data about the educational attainment of Americans in a variety of subject fields. In planning the assessment, the staff quickly came face to face with the problem of defining what the educational system was trying to achieve. In essence the problem was to define the competencies possessed by the good student in areas ranging from arithmetic to citizenship. A tentative list of objectives was determined for each content field by subject matter specialists. These objectives and suggested procedures for gathering evidence about their mastery were submitted for evaluation and revision to panels composed of educators, concerned citizens, and scholars representing various points of view and sections of the country. Some of the panel sessions were heated; some

resulted in objectives being deleted because consensus regarding their appropriateness or importance could not be reached. In the end, however, a general consensus was reached and a list of objectives acceptable to representatives of diverse segments of the populace was advanced in each subject area.

Thus, once a performance-based program has generated an initial specification of the competencies it seeks to instill in prospective teachers, numerous interested and affected publics such as parents, teachers, students, administrators, scholars, and others should have an opportunity to judge, criticize, and revise the competencies. Not only does such a screening process have a public relation value, but more importantly, it affords input into the process by publics who are likely to be vitally concerned and affected by the performance-based approach. It is likely that reconciling the inputs and judgments of varied constituents will be difficult, time consuming, and often frustrating. However, the importance of the specified competencies in directing all aspects of the performance-based program mandates such an evaluation at this early stage.

Evaluating Student Progress

Once some degree of consensus regarding the ends of performance-based teacher education is reached, the problem of evaluating student progress through the program becomes relevant. There are three domains which call for student evaluation. The first involves knowledge of processes, theories, techniques, etc., information which is best evaluated by paper and pencil means similar to those utilized in nonperformance credit and course-centered approaches. The second domain involves teaching practices and activities, demonstrations that the student can actually perform and utilize various strategies, modes, and techniques. Finally, students will be evaluated in terms of the extent to which their knowledge and performance capabilities result in improved classroom learning.

The third domain, which may be termed the product or output domain, will be discussed in a later section of this paper. In this section I will consider the knowledge and practice domains. Three evaluative issues appear relevant to the discussion: the specification of entry behaviors, the mode of grading student performance, and the problem of obtaining an adequate sample of student behavior.

While performance-based advocates stress the exit behaviors of students who pass through their programs, it may be appropriate to raise some questions related to student entry behaviors. Note that emphasis on exist behaviors implies an acceptance of the proposition that enough is known or can be known about instructional techniques to insure that most individuals can attain the prespecified competencies of a program. Despite some personal reservations about such an assumption, I am willing to concede that it may be appropriate and true. However, one might question whether it is worth the time, effort, and expenditure for a particular student to take 8 years to obtain ultimate initial certification when the majority of students reach the same stage in 2 years Should, then, prospective candidates for performance-based programs be evaluated with respect to their attainment of a minimum set of entry competencies before being accepted into the program?

Undoubtedly some initial screening of candidates is always done. Candidates who do not possess adequate academic records or test scores are rejected. Candidates who do not manifest particular interest, personality, and value characteristics are not accepted. But evidence about the cognitive and affective status of individuals is typically gathered by means of global, imprecise, impressionistic means, the very means performance-based programs strive to eradicate. It would appear worthwhile, then, to evaluate prospective students in terms of a set of entry competencies which are specified in the same manner as the exit competencies. Certainly it is impossible to identify a complete list of entrance competencies at this point in time. However, some entrance criteria can be identified and made explicit. Such criteria would require monitoring and revision over successive groups of students to arrive at a more complete and relevant set, but the impact of well defined entry competencies may be worth the effort insofar as they may reduce the frustration of unqualified candidates as well as serving as a means of placing accepted candidates into the optimum instructional starting point (Airasian and Madaus, 1972a).

Once students are into the instructional process, it is necessary to determine whether or not they have attained mastery of the required competencies. The first issue here is, of course, what is meant by mastery. That is, one must identify the standard of performance students must attain to be certified as possessing a competency. Obviously, there is no source, other than judgment, to which one can refer to select appropriate standards. The question of standards is one which plagues all evaluation efforts. However, the nature of the competency, its relevance to instruction, its suspected impact on classroom learning and other such considerations should be weighed in setting the standard. For example, based upon consideration of the

above areas, I have set a mastery standard of 85 percent in my secondary level course in tests and measurements. A student is certified as mastering tests and measurements when he demonstrates competency on 85 percent of my course objectives. Were I teaching physical education to secondary education majors concentrating in areas other than physical education, I would set a lower mastery standard. At any rate, one must have some criterion which enables him to determine whether a given performance is indicative of competency.

With the prior specification of competencies and standards, the problem of grading students is relatively straightforward. The crucial judgment to be made is whether the student has attained mastery of a competency, not how he has performed relative to his peers (Airasian and Madaus, 1972b; Popham, 1971). As various authors concerned with performance-based teacher education have indicated, criterion-referenced as opposed to norm-referenced evaluation is needed. With respect to each required competency, we will want to know whether the student has achieved criterion-level performance, not how high or low he stands relative to other students. Criterion-referenced evaluation, which many argue is more relevant, humane, and "fair" than norm-referenced evaluation, is possible only when criterion performance is predefined. One of the major advantages of performance-based models is that they do incorporate prespecified, well defined exit behaviors. As an added benefit, the availability of clearly stated exit behaviors which serve as evaluative criteria enhances diagnosis of student learning difficulties. Unlike the course grade method of evaluating performance, a criterion-referenced approach related to identified competencies can provide the student with more precise diagnostic direction than "study more" or "work harder." In a sense, all evaluation taking place prior to actual certification of competence is formative, diagnostic evaluation.

Actual data gathering techniques to evaluate knowledge and practice competencies are not complex. For knowledge competencies, paper and pencil tests, oral examinations, and the like are appropriate. For practice competencies, studies' performance in classroom, microteaching, or other similar situations can be evaluated by one, or preferably more, judges on the basis of checklists or overall performance. The basis for judgment, be it judgment of knowledge or practice, is always performance relative to the predefined competencies. The extent to which evidence gathering situations permit students to manifest the behaviors inherent in the competencies is the extent to which the evaluation is valid.

One note of caution must be introduced, however. Any testing situation provides only a sample of a student's behavior. Relatively slight differences in test items or classroom situations from those utilized in an evaluation may engender different student performance. Thus, in judging competency, particularly of teaching practice behaviors where evidence gathering and evaluation are more time consuming, it will be important to include more than a single measurement. Students should be required to demonstrate competence in classrooms of varying types before they are certified as competent.

Competencies Evaluated

In addition to helping judge curricular adequacy and student learning, evaluation plays another powerful role in any instructional system: it defines for those being evaluated what the important or "real" aims of the instructional program are. We have all heard students remark, "Forget what he tells you in class. If you want an A, memorize the book." The implication of such a statement is that the grading process—that which ultimately counts—is based upon a very limited subset of the totality of instructional activities and aims.

The relevance of evaluation's role in defining the real aims of instruction for performance-based teacher education relates primarily to affective performance. No performance-based approach worth its salt will argue that teaching competence is comprised solely of cognitive and psychomotor skills. Attitudes, values, interests, preferences, and the like are at least as important to the competent teacher as cognitive and psychomotor ability—perhaps more so. However, instructional programs based upon the clear identification of objectives or competencies have a lengthy history of paying lip service to the importance of affective outcomes, but concentrating evaluation efforts on more easily measured nonaffective areas. By following such a practice, evaluators define the cognitive and psychomotor aims as the really important objectives of the program, regardless of what the program designers claim about the importance of affective aims. It doesn't take long for students to "psych out" what is required of them and to shape their behavior to comply with expectations.

Thus, if affective performances are important, they must be evaluated. Unquestionably the state of the art of affective assessment lags behind cognitive or psy-

chomotor assessment. In the end, interpretive judgments based upon both formal and informal observations and discussions will probably provide the optimum means of gathering affective evaluative data about student progress. The lack of objectivity associated with such techniques in comparison to more formal paper and pencil techniques should not deter evaluation. One method of stressing the importance of affective aims is to diagnose and evaluate them.

Data Maintenance

In the course of evaluation in performance-based teacher education, much data about individual students will be gathered and assessed. Two issues related to the maintenance of such data must be considered. First, one must consider the form in which individual student data will be kept and disseminated. Second, one must consider to whom data will be disseminated. Each of these issues will be considered in turn.

Performance-based educational programs start with a clear definition of the capabilities possessed by the good teacher. The capabilities are specified in behavioral terms, not in terms of number of credit hours, grades in courses, and the like. It seems reasonable, given differences between performance-based and more traditional programs, that student data within performance-based approaches be maintained in a different manner than in terms of course grades or credits earned. Rather, student records should contain an indication of student attainment of every relevant performance.. While the prior statement may hardly sound insightful or revolutionary, recordkeeping and dissemination in practice are likely to become time-consuming and unwieldy. One of the few advantages of the letter grade or course credit recordkeeping system is that it permits us to describe a student in a short, efficient manner. An index card suffices to characterize a student over 4 college years. In an attempt to overcome the lack of specification inherent in grades and credits, performance-based approaches turned to the identification of specific behaviors related to good teaching. To characterize student learning in the light of the extensive lists of performances identified in most performance-based programs, pages and pages of records are needed for each student. It is important that such records be kept and disseminated in unshortened, unexpurgated form to state certification boards, prospective employers, and to the students themselves.

Individuals responsible for recordkeeping in performance-based approaches must resist the inevitable pressures from school administrators and certification boards to provide a number or a brief statement to characterize a student. To succumb to such entreaties, no matter how attractive they appear or strongly they are urged, will result in reducing the importance of the prespecified performances, which represent after all, the basic appeal of performance-based approaches. Further, to succumb to such entreaties will return certification and hiring practices to a "highest number wins" basis, with the result that specific relevant capabilities will be overlooked or ignored.

Adequate recordkeeping represents only one aspect of the data maintenance problem. To a much greater extent than more traditional approaches, performance-based instructional programs will have on file explicit affective judgments about students. While our schools and society accept cognitive information about individuals as a prime currency of worth or merit, there is some reluctance to judge individuals on the basis of personality, value, or other affective characterisitics. There is a feeling that these are more private and less necessary for others to know. That affective qualities are important to good teaching, few would argue. However, a crucial question to be answered within the context of performance-based programs is "To whom should judgments about a given student's values, personality, interests, and preferences be disseminated?" Clearly some individuals and agencies will require such evaluative information; but which, under what circumstances, in what form, with what guidelines, and for how long are only a few of the concerns which should guide dissemination. There are no right or wrong, ethical or unethical answers to these questions. It is important, however, that every performance-based teacher education approach find answers to these questions, answers which protect the privacy rights of its students yet afford information access to those to whom it is essential.

Judging Teachers by Their Students' Achievement

The final evaluative issue to be raised is undoubtedly the most emotion-charged aspect of performance-based teacher education: the evaluation of teacher competence in terms of student learning. I suppose that in a real sense, my own dual responsibility as a professional educator and a parent reflects the problems involved in judging teachers' performance by their students' learning. As a professional educator, I recognize the myriad of factors outside the teacher's control which affect

learning. But as a parent, I wonder whether not holding teachers even minimally accountable for student learning is an acceptance of the position that the reason for all failure rests upon the student's shoulders.

The data which must be gathered to evaluate teachers' effects upon student learning are not at all clear. Personality and cognitive differences between teachers teaching the same type of classes as often as not result in the same overall class performance for both teacher types. Research indicates that a substantial portion of the variance in student ability and achievement is attributable to environmental factors whose major impact on the student occurs before the student reaches the school. The effect of failure in prior years upon subsequent performance remains the subject of discussion among educators. Although none of us relish the implications, it may be that prolonged failure in school experiences ultimately results in students who are simply not reclaimable, regardless of any teacher's competence or perseverance. We might even question whether failure or poor performance in some courses is not a benefit to students, especially if it prevents them from pursuing careers or areas where they have little probability of success. Finally, and perhaps most importantly, there is the problem of the criteria which indicate successful student learning. The possibilities here are myriad.

In sum, while it is always possible to evaluate teaching competency by measuring student learning, the issues remaining to be settled before such evaluation can be undertaken in an intelligent manner, fair to both teachers and students, suggests that student learning measures not be used to evaluate individual teachers at present. Rather, research on variables and measures hypothesized to relate to student achievement should be undertaken. Particular teacher competencies should first be validated against measures of student behavior. Note, however, that the most accurate and extensive research data will not remove the ultimate need for making evaluative judgments concerning the type and form of student behavior which will be considered indicative of competent teaching.

Conclusion

This paper has argued that the real evaluation questions inherent in performance-based teacher education relate not to constructing objective evidence gathering techniques or measures, but rather to judgments about what data are to be gathered, from whom, and under what conditions. The selection of instruments and techniques is an outgrowth of these prior evaluative decisions. Those whose primary interests are concerned with how questions, need only consult most measurement texts or the curriculum evaluative literature. If I have left some of my audience unsatisfied because I did not address more practical how to questions, my sole defense is that I believe identification and resolution of the what questions is the surest manner of making performance-based teacher education a more viable, effective, and humane approach to the problem of educating teachers.

References

Airasian, P. W. and Madaus, G. F., Criterion-referenced testing in the classroom, NCME Measurement in Education, 3, 4, May, 1972a.

Airasian, P. W. and Madaus, G. F., Functional types of student evaluation, Measurement and Evaluation in Guidance, 4, 4, 1972b, pp. 221-233.

Airasian, P. W., Madaus, G. F., and Rakow, E. A., An instructional evaluation monitoring system, Curriculum Theory Network, Monograph Supplement, Curriculum Evaluation: Potential and Reality, 8/9, 1972, pp. 74-95.

Atkin, J. M., Using behaviorally stated objectives for designing the curriculum: a cautionary note. Paper delivered at the annual meeting of the American Educational Research Association, Chicago, Ill., February 1968.

Block, J. H., Mastery learning: theory and practice, New York: Holt, Rinehart & Winston, 1971.

Bloom, B. S., Hastings, J. T., and Madaus, G. F., Handbook of formative and summative evaluation of student learning, New York: McGraw-Hill, 1971.

Campbell, D. T., Reforms as experiments, American Psychologist, 24, 4, 1969, pp. 409-429.

Combs, A. W., Some basic concepts for teacher education, The Journal of Teacher Education, 23, 3, 1972, pp. 286-290.

Doll, W. E., A methodology of experience: an alternative to behavioral objectives. Paper delivered at the annual meeting of the American Educational Research Association, New York, N. Y., February 1971.

Eisner, E. W., Educational objectives: help or hindrance. Paper delivered at the annual meeting of the American Educational Research Association, Chicago, Ill., February 1966.

Good, T. L. and Brophy, J. E., Looking in classrooms, New York: Harper and Row, 1973.

Jackson, P. W., Life in classrooms, New York: Holt, Rinehart & Winston, 1968.

Popham, W. J., Probing the validity of arguments against behavioral goals. Symposium presentation at the annual meeting of the American Educational Research Association, Chicago, Ill., February 1968.

Popham, W. J. (ed.), Criterion-referenced measurement, Englewood Cliffs, N.J.: Educational Technology Publications, 1971.

Provus, M. M., Evaluation of ongoing programs in the public school system. In R. Tyler (ed.), Educational evaluation: new roles, new means, sixty-eighth yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1969, pp. 242-283.

Stufflebeam, D. I., *et al.*, Educational Evaluation and decision making, Itaska, Ill.: F. E. Peacock Publishers, 1971.

Tyler, R. W., Basic principles of curriculum and instruction, Chicago: University of Chicago Press, 1950.

Tyler, R. W., Constructing achievement tests, Columbus, Ohio: Ohio State University Press, 1934.

The State Of The Art In Performance Assessment Of Teaching Competence

By

FREDERICK J. McDONALD

To describe and evaluate the state of the art in performance assessment of teaching competence, we should answer three questions: (1) What are the goals of the systems to measure teaching competence? (2) What progress have we made in developing the procedures required for systems measuring teaching competence? (3) What problems must be solved to create effective and efficient systems for assessing teaching competence?

Before attempting to answer these questions, I should first openly acknowledge the assumptions that influence my ideas and those that do not. I assume that even small progress made in assessing teacher competence will be a great improvement over our present evaluations. Because I assume this, I am willing to urge the use of procedures and systems which at the present time are limited or even defective—since I also assume that as we use what instruments and techniques are now available we shall learn more about the nature of teaching competence and progressively improve our methods for evaluating it. I further assume that the cost of measuring teacher competence will not necessarily be prohibitive, despite the fact that many people seem to assume the contrary. It is obvious that instituting a measurement system where there was none before will impose heavy costs. But until we have a complete cost/benefit analysis, I am not willing to assume that these costs will be unmanageable. Further, I do not assume that the evaluation system will necessarily be complex because we do not know as yet what information is needed to make reliable and valid judgments about teaching competence. I choose, therefore, to be optimistic about both the cost and complexity of an evaluation system

It seems to me that the most reasonable position to take at this time is to acknowledge the many problems and to regard their solution as dependent upon the ordinary processes of research and development. It is self-defeating and destructive to so negatively prejudge the situation that work which might resolve the problems is either impeded or prematurely terminated.

The Goals of an Evaluation System

Let us now return to the three questions with which I began this presentation. The first is obviously the most important: what should be the goals of a performance assessment system? The ultimate goal of such a system is to provide information that enables the decision makers—teacher educators, certification officers, and administrators—to decide whether a teacher candidate has sufficient competence to be permitted to teach. Everything else that we will want to say about this goal is either a refinement of what is meant by making a judgment about competence or a problem that must be solved if such judgments are to be made. I will therefore list a set of conditions that must be satisfied for this goal to be attained.

First, measurement procedures used in the evaluation of teaching competence must have high validity. Validity in this case means we have demonstrated a substantial relationship between a teaching skill or performance and its effects upon students.

At this point in time we know very little about what skills or performances have demonstrable effects on student behavior. The highest priority in a research program on teaching competence must be given to solving this problem. The conclusion, however, should not be drawn that we must defer the development of an evaluation system until all the relevant research has been done. We already have an abundance of ideas on pertinent teaching competencies which we can begin to measure—a necessary first step, and one whose effect on teaching performance can be studied systematically as part of the process of developing evaluation systems.

We cannot or ought not, however, make definitive judgments about teaching competence until the validities of specific teaching skills or performances have been established.

Second, we must establish the reliability of assessment procedures that will be used in evaluation systems. Reliability in this case means the stability of a teaching skill or performance. Does the teaching skill vary considerably from day to day, class to class, topic to topic? We need to know the conditions under which it is most likely to be variable and those under which it is most likely to be constant.

While considerable attention is given to the problem of validity, practically no attention has been given to the performance characteristics of teaching skills. There is some intuitive understanding that a variety of conditions may affect a teacher's skill, but formal research on the problem is nonexistent. It makes no difference how research on this aspect of teaching performance comes out because we need all kinds of information in making judgments about competence. If, for example, a particular teaching skill requires special conditions to elicit it, we must ask whether those conditions were present when the teacher's possession of the skill was being evaluated.

Third, we need to know what, for lack of a better term, I will call the "learnability" of a skill. It is conceivable that a particular skill, even though it has been demonstrated to have highly desirable effects upon children's learning, can be learned by only a few individuals. We do not expect everyone to become theoretical physicists or concert pianists or gold medal-winning athletes because we know that some special aptitudes may be required for achievements of this order. Similarly, there may be forms of teaching whose acquisition demands unusual aptitudes. Thus, we should avoid the trap of assuming that every skill demonstrated to have desirable effects on students will necessarily be required of every person who wishes to teach.

We must develop the kinds of information described in these three points—relating to the reliability, validity, and learnability of teaching skills. We cannot achieve our goals unless we can define what is to be measured, under what conditions it is to be measured, and how the information yielded by the procedure is to be interpreted. But there are some other conditions beyond these three that must be met by any system for measuring teaching competence.

A fourth condition, for example, is that the information gathered about teachers must be as uncontaminated by subjective biases and political processes as is humanly possible. Any observation of human behavior is, of course, influenced to a degree by the characteristics of the observer and by the characteristics of the measurement procedure itself. We shall need to develop new strategies for collecting information on teaching competence if we are to minimize the problems of gathering such information.

Fifth, the conditions of measurement must provide comparable information on groups of teachers. We must standardize the conditions under which teacher behavior is measured. Otherwise, it is impossible to tell whether variations in teacher behavior are due to situational differences or to differences in teacher competence, thus exposing the candidate to the vagaries of idiosyncratic favoritism.

The conclusion should not be drawn, however, that the assessment system should ignore variations in the conditions under which teacher performance occurs. As I stated earlier, it is essential that we determine those conditions under which a teaching skill may be expected to vary. Such information is necessary in evaluating teaching competence. But we cannot treat teaching as if it were so different on each separate occasion that we can never evaluate it. The conflict between establishing reasonable expectations for teaching performance and the variety and complexity of the situations in which teaching occurs is one of the most important problems we have to solve. Until it is solved, our decisions about competence must necessarily be tentative.

Sixth, we must develop an evaluation system that makes due allowance for the teachers' opportunities to acquire the relevant teaching skills. A decade may well pass before teacher education programs have become performance-based and before a reasonable amount of performance assessment has taken place. To demand competence of individuals who have not had the opportunitites to acquire that competence is immoral, and to make judgments that affect their careers on such an insubstantial basis is equally immoral. Assessment should *not* be used for evaluative purposes when the individuals being assessed have not had the opportunity to learn those things for which they are being evaluated.

In listing these constraints, I have outlined the major problems that must be solved if a system of assessing teaching competence is to be developed and used in a practical way; indeed the solution of the problems depends on our meeting these conditions.

The State of Measurement Technology and Knowledge

Despite popular opinion to the contrary, measurement technology follows the wants of its users. For three-quarters of a century, decision makers of one kind or another have wanted to assess what candidates for teaching positions *know*. Measurement technology for assessing academic knowledge thus became highly developed. Consequently, we now have widely available tests of knowledge of subject matter and of knowledge about teaching methods. But until recent years there has been no demand of any consequence for the development of performance assessment. As a consequence, this aspect of measurement is largely undeveloped.

We have developed observation scales which permit us to report what teachers do under ordinary teaching conditions; but none of these observation schedules or methods meet any of the constraints which I outlined previously. The conclusion should not, however, be drawn that the use of observation schedules should be terminated; nevertheless, they and any other measurement procedure must be subjected to the kinds of research and development necessary to make them useful within the stipulated constraints.

The basic problem of performance assessment technology is to create a controlled situation or to use an ordinary teaching situation with sufficient knowledge of its characteristics so that their influence on the performance can be assessed. Observation schedules fall into this second group of procedures. The problem of assessing the influences in a wide variety of teaching conditions and performances is enormous. I have come to the conclusion that a more profitable line of research should use controls, simulated teaching situations.

Rather than describe the adequacy of available measurement technology, a rather depressing picture, I shall describe the work in progress on the assessment of performance using simulated teaching conditions. This work is relatively new in the field of teaching competence, and it is conceivable that it may never develop into a practicable technology. But some of the problems must be solved and are more likely to be solved efficiently and satisfactorily by using simulated teaching conditions. The relationships between teaching performance and student learning, for example, can be studied very efficiently by using microteaching sessions. Similarly, the reliability of teaching performances under controlled conditions can also be easily assessed.

We are attempting to build a system that provides

Frederick J. McDonald

information about basic teaching skills, teaching performances, and teaching strategies. This system is designed to use simulated teaching situations of varying degrees of complexity. Initially, both the teaching skill measured and the conditions under which it is measured are relatively simple. The prospective teacher moves through the system from this simple initial level to one where he or she must use a variety of skills, performances, and strategies in combination and must enact them over a period of time. The critical feature of the component systems is that the conditions are carefully controlled. In what ways are they controlled? For any given component, the teaching objectives and teaching performances are specified; for example, if we are attempting to assess the teacher's ability to reinforce students participating in a discussion, the objective of the lesson is set as increasing the amount of student participation in the class. The trainee is told that he will be evaluated on the degree to which students participate in a class discussion. Initially, we want to measure only his use of reinforcement procedures to facilitate participation in classroom discussion. In other components and at other times, the teaching situation will be constructed to evaluate the trainee's ability to present a stimulating problem for discussion or to ask questions which elicit comments from students or facilitate discussion by highlighting differences of opinion.

In order to make assessments at this time, certain factors have to be controlled. For example, if we are to

study the teacher's use of reinforcement techniques, we do not want the teacher preoccupied with the preparation or development of the topic or with the choice of the topic, nor do we want a group of teachers to have different objectives. Therefore, as I noted, we stipulate the objective, and we also provide the teaching materials for the lesson. Further, we control variability of the students by reassigning the students to a teaching assignment and by providing the preparation period for familiarizing themselves with the material to be discussed. It should be noted, and I wish to emphasize, that the factors which are controlled in one type of situation may be systematically varied in another in order to assess the degree of the teacher's skill.

What can be learned by using this technology? These microteaching sessions can be used to assess a teacher's skill under a variety of different teaching conditions. We have tried several formats of microteaching lessons to determine what can be learned by having teachers teach different lessons and different groups of students. In one format, for example, we asked the teacher to teach four different topics to the same group of students; in another, a teacher teaches different students the same topic; and in yet another, the teacher will teach different topics to the same students and then the same topics to another set of students. In these ways we can test how the teacher's skills change under the different teaching conditions. We have already found that inexperienced teachers' behavior varies from topic to topic, to some extent, even though the topics appear to be highly similar in nature and would seem to offer similar opportunities for using teaching skills.

Eventually, we would expect to have a series of representative teaching situations in which it had been shown that certain kinds of teaching skills are highly likely to be useful. An assessment battery would consist of a series of microteaching sessions covering these teaching situations.

Because the microteaching format has several limitations, such as relatively short lessons and small numbers of students used, we developed a mini-course format to use for more complex teaching situations. A mini-course in this context is a short course lasting 1 or 2 weeks. The teachers teach this course an hour a day for one of these periods of time. For these courses we have also developed a complete statement of objectives, course materials, and achievement tests to be taken by the students at the end of the course. Almost 300 preservice teachers have taught these courses so that we have gathered considerable data on teaching performance and experience with the technology itself.

The results of the analyses of these teaching performances indicate that the microteaching performances are relatively poor predictors of the teaching performances in the mini-courses. But, for several reasons, I am not willing to offer this conclusion as solidly substantiated. First, the subject-matter of the short courses is more complex, and segments of any one lesson would be similar to the teaching situations created by the microteaching session. But their combination presents a different kind of teaching problem. Also, in the situation in which we have gathered data, relatively little performance training took place between the initial microteaching and the subsequent teaching in the short courses. Systematic training using the assessment gathered in the microteaching might influence substantially the relationships between the two kinds of performances; that is, where the teacher has sufficient skills, as revealed in the microteaching performance, the instructors would reinforce these skills so that they were more likely to be used by trainees in the more complex teaching. There is still reason to believe that we have created two sufficiently distinctive types of performance situations which will yield equally valuable but different kinds of information. My own opinion is that the microteaching is more useful for assessing the degree to which a teacher has basic skills; the mini-course is most useful in assessing how teachers integrate these skills into complex teaching performances.

As I said earlier, our goal is to build a system that will assess the simple components of teaching competence and also its more complex aspects. We visualize the coordinated assessment system as made up of periods of microteaching used primarily for assessment purposes and short courses or units that can later be embedded in internship and student-teaching experiences. These components would be arranged so that the training institution could be given information continuously about the level of competence being reached by the trainee.

The student teaching and internship experiences can be used to assess daily performance under uncontrolled conditions. They are useful for providing information on what teachers are likely to do in contrast to what they are able to do. The simulation-type teaching situations that I have described should be thought of as teachers assessing what teachers are *capable* of doing; whereas on-the-job performance tells us what he or she regularly does. We can also use on-the-job observation to assess such factors as teaching style.

I am convinced, after a year's experience with ad-

ministering systems such as these and another year's analysis of the data gathered, that it is entirely feasible to build a useful performance assessment system that can be embedded in any teacher-training program. As such systems are developed, state departments of education can learn what criteria ought to be applied to the assessment of teaching competence. The prototype that we are developing in our research program can be used to generate concepts about the appropriate means to evaluate each training institution's procedures for assessing the competence of their students.

It is obvious to me that the type of techniques being developed can be used with beginning teachers prior to the granting of tenure, and for periodic reassessment of teaching skill. We should not, however, use these procedures at the present time for continuing certification schemes, any more than we ought to use the current observational procedures and administrator evaluations for that purpose; but there is good reason to believe that, in time, assessment strategies can be developed which will be useful both in preservice and inservice assessment of teaching competence.

One question asked frequently is, "How expensive will performance assessment be?" As I observed earlier, there is no adequate way not to answer this question because we do not know what information we shall need to make judgments about teaching competence. My experience indicates that we can gather large amounts of data about teaching performance relatively inexpensively: for example, three 20-minute microteaching sessions yield an hour of teaching performances. A videotape of these performances can be observed to assess many different skills. In fact, the short course yields almost more information than can be quickly processed. We have resorted to audiotaping and videotaping about a half hour of each hour session for the very practical reason that this half hour covers three-fourths of the teaching that occurs. I am not convinced that even this much information is needed. We are studying this problem by taking different samples of the videotape performances that we have and finding the intercorrelations among teaching performances across the samples. This technical work will eventually help us find the optimum amount of observation time that we must use in order to make reliable judgments about teaching skills.

While many people are concerned about the cost of performance testing, few discuss the economies that may be achieved by using them. It is very apparent that, when one looks at videotapes of teaching performances, a diagnosis of basic teaching skills can be made very easily, and this information can be used to speed the learning of these teaching skills. We already have sufficient information on feedback and modeling techniques to know that the information generated in a performance assessment system can be incorporated into feedback and modeling systems to increase the effectiveness of training. Thus, we may expect the institution to benefit by achieving economies in training time and the trainee to benefit by achieving competence sooner when he enters upon actual teaching.

Reactive Simulation Devices

I have been describing performance assessment procedures which require the teacher to interact with students in a simulation of actual teaching. There are some other procedures that appear to be promising in assessing potential teaching competence. We have developed a filmed simulation test that portrays a teacher conducting a class. The film is stopped periodically and the viewer is asked to say what he or she would do in this situation; in other parts of the test the viewer is asked to explain what is occurring in the class, and, in some places, he is asked what advice or suggestions he would give to the teacher. We do not yet have data on the predictive validity of this measurement, but it obviously taps the prospective teacher's attitudes toward students, his perceptions of the problems of teaching, and his understanding of some phases of teaching strategies. The test situation is sufficiently lifelike to create great interest on the part of those who view the film, and invariably the film produces some intense discussions about what the teaching portrays, thus suggesting to us that the film also has great potential for training purposes.

Another type of simulation device has been created to depict the kinds of strategies teachers use. This procedure is a gamelike device in which an experimenter plays the role of student in a highly controlled way. The teacher arranges the subject matter in the form of presentations or questions, and the experimenter responds whenever the teacher asks a question. This gamelike situation does discriminate sharply between deductive and inductive teaching styles. It seems likely that the role of the experimenter can be computerized so that the game can be used by large numbers of trainees in a more efficient manner.

Both of these techniques are illustrative of the kinds that are likely to be developed to assess teacher competence. Obviously, their validity in predicting teaching performance has to be established. But the work to

date suggests that they are probably of sufficient validity to be useful in assessing what a teacher is likely to do in a classroom.

The Problems To Be Solved

The work that I have described here is meant to be illustrative of the possibilities of assessing teaching competence. I emphasized that the problems are primarily technical and partially theoretical. We have not yet developed an adequate technology of performance assessment, but the reason for this is related more to the lack of experience with these technologies than to a lack of knowledge of what to do.

The theoretical problems are more complex and will require extensive research and development. The kinds of procedures that I have been talking about here can be adapted to experiments—for example, assessing the relationship of teaching skills to pupil learning. Carefully controlled experiments can be designed in the microteaching format to study hypotheses about these relationships. The mini-course data can be used to induce hypotheses about the relationship of teaching competence to pupil learning. But until the basic questions about what competencies make a difference have been resolved, the design of assessment technology will be incomplete.

Each of the conditions governing an assessment system, that I outlined previously, presents a technical problem to be solved. These problems are highly amenable to research, and there is no reason to believe that they cannot be solved in a relatively short period of time.

A basic problem is political in character. Many individuals believe that competence cannot be measured because they do not want competence to be measured. The solution to this problem is obviously complex. Certainly the development of efficient and demonstrably useful assessment systems will weaken the arguments against assessment. Still, there will probably always be those who will say that the most important aspects of teaching are unmeasurable. But I, for one, have never been able to understand how this conclusion can be reached before anybody has tried seriously to measure the more elusive aspects of teaching.

Conclusion

What should state departments of education, who wish to move to performance-based certification, do about the assessment problem? Obviously, an enormous amount of research and development work remains to be done, and even though we are optimistic about the pace at which the research and development can proceed, we cannot at this time offer state department personnel finished performance assessment systems. For that matter, we all know that just selecting the competencies that will be the criterion performances cannot be done at the present time.

I have the impression that state department personnel are largely concerned, and justifiably so, about this latter question. I propose that rather than think about assessment in terms of what competencies should be measured, state department personnel think about the criteria for evaluating assessment systems from which judgments about competence are made. If you agree with me that it will be several years before we have a good body of research established on the relationship of teaching competence to pupil learning, then it seems to me that the most practical course to take is to set standards for the evidence about competence that is to be provided, however that competence may be defined.

I urge your consideration of this approach because I believe that once a faculty in an institution becomes concerned with the assessment problem, the definition of competence and the demonstration of its relevance to student learning will become major concerns of the faculty and will more likely be tackled quickly and efficiently. But it also seems to me that what state department personnel ought to be asking is what the bases are to be on judgments about competence are made. If attention is focused on these decision-making processes and the information is used to make decisions about competence, we shall at least have guaranteed that the judgments are not being made subjectively or that a multiplicity of factors affecting competence have been inadequately controlled or that the range of a teacher's skill has not been appropriately evaluated. The problem for the next several years can be stated as a question addressed to the training institution, "Given your conception of what constitutes competence, what evidence have you gathered that demonstrates that teachers have acquired these competencies?"

In my opinion, if state departments would adopt criteria for the evidence to be presented to support teacher certification, the performance-based movement would move forward quickly. You would also avoid the trap of falling back into approving programs which have a performance base but cannot produce evidence that the trainees who go through the program are in fact competent teachers. This approach would focus the institution's attention on the development of the as-

sessment of competence and would give each institution greater freedom in the development of programs.

Those of us who are interested in the performance-based movement in teacher education accept a basic principle: we do not care how a teacher becomes competent provided that he or she can demonstrate that competence. If we believe in this principle, as I know we do, then our efforts should be directed to deciding what evidence we will assess to determine whether a teacher is competent.

Conference Participants

PBTE: Proceed With Caution

By

BARAK ROSENSHINE

As I review the current state of knowledge about PBTE, I am overwhelmed by the certainty expressed by state legislatures and state departments of education when they mandate teacher performance criteria as an answer to educational problems. It is equally overwhelming to read the lists of behavioral competencies developed by teacher educators or to read the training packages developed to instill these competencies. Do these educators and legislators know something that I don't know, and why hasn't this message gotten through to me?

When I look at research on teaching, I am overwhelmed with uncertainties. This is because systematic, cumulative research on teaching behaviors and student learning has barely begun. The results of the research, to date, are best seen as providing suggestions for future research, not future practice. Research to date tells us very little on effective questioning skills, for example, and even less about whether effective questioning skills are similar across grade levels and subject areas.

Until informed otherwise, I assert that it is illusionary to mandate teacher performance criteria and expect that pupils will benefit from teachers achieving a set criteria on these skills. We simply know too little, and our research is too contradictory to support such mandates. Therefore, I would caution leaders in PBTE to proceed slowly and with caution. Teaching competencies are not matters that can be decided by a poll of concerned citizens or be legislated by state legislators.

At the same time, the interest generated by PBTE, and the training which is being provided under PBTE programs both provide settings and possibilities to conduct research to answer some of the questions which are becoming increasingly important. The purpose of this paper, then, is to attempt to specify research problems and suggest ways in which such research can be conducted.

Current Knowledge on Teaching Competencies

Anyone who writes about teaching competencies is aware of the lack of research which links teaching behaviors to beneficial effects upon pupils. My own view is that it is premature to use this research to derive implications for practice or teacher training. This is because there have been relatively few studies of teaching behavior and student learning, and most of these studies were done on a small budget with a small sample of teachers and a small number of observations and outcome measures. At present, the conclusions *are not ready* for translation into teacher training competencies. At present, we have only the faith that increased research, and *improved* research will yield knowledge that can be translated into teacher competencies.

The following is my encapsulated view on the state of research on teaching competencies and student growth:

a. The research base for building teaching competencies is extremely thin because there have been only a small number of studies which attempted to relate teaching behaviors to pupil learning and very few reviews of these studies. For example, although one extensive review was written on the correlations between teaching behaviors and pupils' cognitive gain (Rosenshine, 1971), similar reviews apparently have not appeared on the correlations between teaching behaviors and pupil social or cognitive growth. Reviews on *experimental* studies on teaching behaviors and pupil learning have yet to appear.

   Because of the lack of research cited above and the unevenness of the quality of research which has been done, the results of the best of these studies are not sufficiently clear to be translated into performance modules.

b. The correlational studies which have been completed and reviewed (see Rosenshine, 1971; Rosenshine and Furst, 1971) are best seen as providing ideas for future research, not future practice. In each of the reviews which I authored or coauthored there is no section on implications

for practice. Rather, most of the space is devoted to ideas which might be considered in future research studies. In the most comprehensive of these reviews (Rosenshine, 1971) the presentation of the results on each variable is followed by suggestions for future research. These suggestions are made not only for results which were consistently statistically significant or consistent in trend, but also for results which are, to date, nonsignificant. Thus, investigators who wish to build upon current research will have no trouble finding suggestions for such work. Whether these suggestions have merit is still to be decided.

e. Although a variety of teacher-training skills packages have been produced by educational laboratories, research and development centers, private industry, and private individuals, few, if any of these packages are accompanied by technical reports which show that training teachers in these skills (or a combination of skills) results in greater student learning.

The above paragraph is particularly distressing when one learns that many of these training materials have been available for 5 years or more, and thousands of preservice and inservice teachers have received training in these skills. Yet, when one looks for reports on the effectiveness of such training for helping pupils growth, there is a lack of reports. Those few reports which are available were usually dissertations which focused upon 15 to 20 teachers in one location.

This lack of research on products developed by agencies other than colleges of education is distressing for a number of reasons. First, it perpetuates a familiar educational problem of develop and disseminate and let validity lie in the eye of the beholder. If well-funded organizations have not adequately tested their products, what can be expected from colleges of education? Second, the lack of research on pupil growth deprives the developer of vital information which can be used in modifying the product or advising how the product can best be used. Third, it leaves colleges of education in a quandry when they have to select appropriate training programs.

d. The lack of research in the past makes it very difficult to design research for the future because we have little to build on. Currently, when experimental studies are conducted and nonsignificant results are obtained, there are any number of explanations: the design may have been inappropriate, the tests may have been inadequate, the treatment itself was not very important, or any of a number of additional possibilities. Without a research base to which we can refer, we can barely begin to guess which alternative explanation is most likely.

Will Research Take Place in PBTE?

Performance-based teacher education offers an excellent opportunity for colleges of education to aid in expanding and clarifying our *knowledge base* about teaching competencies and student growth. A critical question, however, is whether there will be an extensive research effort or whether the resources will be devoted to development and a vague promise for research at some future time.

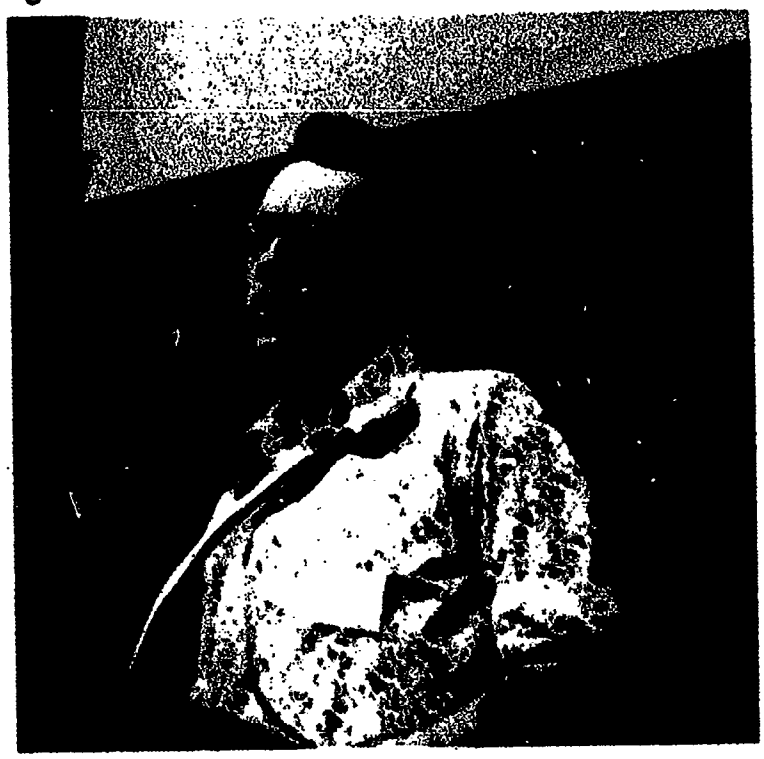
This section is written in the hopes of facilitating such research. It begins with a listing of existing teacher-training materials, raises issues about the selection of teaching competencies, discusses those agencies which might conduct such research, and concludes with a few suggestions on research settings and outcome measures.

Sources for Teacher Training Products

As I survey current work in PBTE, it seems that a great deal of time is being given to developing training materials. I question whether this time is necessary. In my opinion, *we already have a large number of developed products* which can be used either in present form or with slight modification in PBTE. Some of these products can be identified fairly easily, and current work will identify even more.

A major source for identifying training products is "Resources for Performance-Based Education" (Houston, 1973), a joint product of the New York State Education Department and the Multi-State Consortium on PBTE. The book is the most comprehensive compilation of materials I have seen. It includes over 3,000 teacher-training products which were developed by private individuals, universities, publishing houses, and government-sponsored organizations. Another source of teacher training materials is the third edition of the CEDaR Catalogue (1972). This two-volume edition contains descriptions of products which have been developed by the federally sponsored educational laboratories and research and development centers. It also contains descriptions of products under development,

and elaborate cross-references of the products. The CEDaR Catalogue, however, is currently limited to the products of the 20 organzations in the Council for Educational Development and Research (CEDaR). In addition, the Program on Teaching Effectiveness at the Stanford Research and Development Center has been collecting information on teacher-training products developed by all sources and a publication of their findings should be available in late 1973.



Barak Rosenshine

If such a large number of teacher-training products are already available it would seem unwise to develop still more. A more efficient strategy would be to use or modify these to meet institutional needs. An example of such a program is the one developed at Florida International University (Sobol, 1973) in which all of the modules represent modifications of existing, accessible material.

Because there already exist a large number of teacher-training products, and because we have little information on the effectiveness of these products for enhancing pupil growth, it seems reasonable to advocate that research on existing products take a higher priority than the development of still more unvalidated products.

Yet, the evidence to date strongly suggests that still more development is outpacing research in PBTE. This is not surprising because most educational developments of the last decade have been characterized by "implement and develop on a large scale and *then* do research." Unfortunately, the research is seldom done, and by the time we are ready to do serious research, the developers and their public have become occupied with still another "new" but unvalidated innovation and return to more development with another promise of research at some future time.

Issues in Selecting Training Products

The availability of a large number of different teacher-training products (there appear to be over 1,000 modules in "Resources for Performance-Based Education," and a larger number of audio tapes, programed texts, kits, games, and guides) makes the problem of selection extremely difficult. Some questions and issues in selection might be framed as follows:

1. How does one select teacher competencies?
2. How does one distinguish between competencies which are useful and those which are trivial and misleading?
3. Are these competencies appropriate for *all* subject areas, *all* age levels, and *all* types of students? If not, how does one decide the most appropriate competencies for the particular setting?
4. Which competencies are most appropriate for different contexts (e.g., types of student, differing school environments, differing instructional materials, differing grade levels, and geographical areas)?
5. Are the same competencies useful for cognitive, social, and emotional growth in pupils? If not, how does one decide the most appropriate blend?

The common practice for answering these questions has been for people regarded as knowledgeable in the field to use their experience and knowledge to make a best guess. Because there is a lack of research, the guesses represent the best thing we have. The critical question then, is whether there will be future, systematic, cumulative efforts to attempt to answer the above issues.

Who Will Do the Research?

One obvious locus for research is those institutions doing the development, particularly the educational laboratories. My guess would be that if educational laboratories are also involved in research on the effectiveness of their teacher-training products upon pupils, the results of this research would feed back and improve the development of present and future materials.

But the major place for research on the teacher-training performance-based packages would be in those institutions which are preparing teachers. Those institu-

tions preparing 500 to 1,000 teachers a year would have more than a cursory interest in the effects of their offerings, not only upon their trainees, but upon the pupils which the trainees will be teaching.

Collaborative research would seem appropriate, particularly, within a state or across a national organization such as the Teacher Corps. Within the pages of the PBTE newsletter there are a least two examples of proposed collaborative research. In one, involving Teacher Corps trainees, a study is planned for the 1973-74 school year involving 400 interns in 20 institutions. In this study, approximately 10 measures of cognitive, emotional, and social growth of children taught by these interns will be collected in the fall and again in the spring. For each intern, information will be collected on the extent to which they possess the skills, attitudes, and knowledge that are believed to facilitate the learning and growth of minority group and low-income elementary school children. A major research question will be the relationship of intern skills, attitudes, and knowledge to the learning and growth of their pupils.

Another program described in the PBTE newsletter presents a two-step process. First, statewide institutions are to implement a PBTE program which includes assessment of whether trainees achieve desired skills, knowledge, and attitudes. In the second step, the institutions are required to demonstrate whether the desired skills, knowledge, and attitudes are appropriate. It is expected that both steps will be operative before 1980.

Funding PBTE Research

The funds for the proposed research can come from many sources. One source should certainly be those legislatures which are advocating teacher certification based upon performance. Another source could be those educational institutions which plan to develop performance-based training modules. Even if modules weren't available, one would expect that for every $2 spent on development and dissemination, $1 would be spent on research to validate the training behaviors against desirable educational outcomes. Another source could be the Federal Govenment, which, to my mind, has not been allocating a sufficient proportion of funds for research into the developments they are funding.

Settings for Research on Performance-Based Skills

The appropriate settings for research on performance-based skills is a tricky question which probably will not be answerable until a good deal of research has been done. The major issue is the extent to which results obtained in special settings are generalizable to regular school settings. Special settings have been used in a number of studies, and these range from 10-minute lessons which student teachers present to specially assembled groups of pupils to unique 2-week units which regular teachers present to their regular pupils. The differences in settings, instructional materials, and outcome measures are such that research will be needed to determine the extent to which results obtained in one setting are generalizable to another.

At present, a case can be made for the utility of all settings. Small controlled settings offer possibilities for intensive study, and findings obtained in these settings can be verified in more natural situations. Thus, although teacher-training institutions would have difficulty conducting research in regular classrooms, they could conduct research by assembling smaller groups of pupils and noting the effects of trained and untrained preservice teachers who taught mini-lessons lasting from 10 minutes to five 1-hour lessons.

To date, the number of such studies which have taken place on the preservice level is extremely small —less than one per state. It is hoped that colleges of education will recognize the advantage of combining the implementation of PBTE with research on the performance-based skills.

Matching Outcome Measures to Training

One of the biggest problems in conducting research on teaching, whatever the setting or the performance skills, is selecting the appropriate criterion measures. In many studies, the evidence is limited to whether *trainees* modified their instructional behavior in desired ways. Research which contains trainee outcomes is important, but it cannot substitute for research on whether these trainee outcomes increase pupil learning.

If someone is conducting validation research, it presently seems extremely important to spend a great deal of time inspecting the match between the teacher performance skills and the criterion measures. Such inspection and matching can take place two ways. One way would be to start with specific performance skills (e.g., asking higher order questions) and make reasoned guesses on the kind of pupil learning one can reasonably expect if these specific skills are implemented. (One might also consider the student learning which would not be expected.) Then one could locate or develop appropriate learning measures.

A second procedure would be to select pupil outcomes (e.g., social or emotional growth), inspect the available measures in this area, and then develop a rationale for the performance skills which seem appropriate for enhancing such pupil learning. This second procedure has been used to reanalyze correlational and experimental studies in which cognitive measures are the outcome. The current evidence suggests a fairly high specificity of effect. That is, there is a strong correspondence between the cognitive activites emphasized during instruction and those assessed on the cognitive outcome measure. If further analyses continue to support specificity of effect for cognitive measures, such specificity may also hold for affective outcomes. If so, then it would seem wise to inspect the affective outcome measures used in a validation study to see if there is sufficient correspondence between the affective behaviors stressed in the instructional modules and those assessed on the outcome measures.

Currently, we spend relatively little time inspecting the match between the performance skills and the outcome measures. Rather, the more common practice has been to select performance skills from a variety of sources, train teachers in these skills, and run some experiment to see whether the pupils of the trained teachers show greater learning than the pupils of the untrained teachers. The usual result has been no significant differences.

(The reader who is interested in further study of the design of validation studies might consult Flanders (1970, chapter 12), Flanders (1971), and Rosenshine and Furst (1973).)

Irrefutable Hypotheses

One hopes that when validation research occurs in PBTE, investigators and reviewers will allow for the possibility that certain pet ideas will *not* be validated. Unfortunately, there has been a tendency in education to hang on to all ideas, no matter what the research results. When faced with studies which do not support our pet ideas, we frequently claim that if the studies had been better designed, or if the treatment had lasted longer, *then* our ideas would have been validated. When all else fails, we invoke The Educators' Creed:

> These are things that tests can't measure
> These may be the most important things of all
> And the experimental group did best on these.

When training teachers in performance skills does not yield differences in pupil growth there are any number of possible explanations, and these explanations are sources for future research. It seems unreasonable and unproductive to develop lines of argument which allow us to assert the validity of all performance skills which we consider important no matter what empirical results are obtained. Indeed, given the large number of teaching skills which have generated, one would hope that research would help us *reduce* the number of skills currently considered important.

Summary

In performance-based teacher education (as in other areas of education such as the teaching of reading), we are faced with a large number of training and instructional materials, a great deal of interest, and a limited amount of research. As such, PBTE appears no better and no worse than the rest of the educational field.

Whether PBTE will be limited to development and dissemination, or whether it will also include necessary research on pupil learning is a critical question which will be answered by actions in the next few years. The teacher-training materials which have been developed offer excellent opportunities for experimental research (as contrasted to the correlational research which has dominated the field to date.) One hopes that institutions engaged in training teachers will cooperate in developing a systematic and cumulative knowledge base.

References

CEDaR Catalogue (3d edition). Commercial-Educational Distributing Services, P.O. Box 3711, Portland, Oregon. 1972.

Flanders, N. A. "Analyzing Teaching Behavior." Boston: Addison-Wesley, 1970.

Flanders, N. A. A national coordinated program of research on teaching effectiveness. In "How Teachers Make a Difference." U.S. Government Printing Office, Washington, D.C. Stock No. 1780-0813. 1971.

Houston, W. R. and others. "Resources for Performance-Based Education." The State Education Department, Division of Teacher Education and Certification. Albany, N.Y. 1973.

Rosenshine, B. "Teaching Behaviors and Student Achievement." New York City: Humanities Press, 1971.

Rosenshine, B. and Furst, N. Research on teacher performance criteria. In B. O. Smith (Ed.) "Research in Teacher Education: A Symposium." Englewood Cliffs, N.J.: Prentice-Hall, 1971. pp. 37-72.

Rosenshine, B. and Furst, N. The use of direct observation to study teaching. In R. M. W. Travers (Ed.) "Second Handbook of Research on Teaching." Chicago: Rand McNally and Co., 1973. pp. 122-184.

Sobol, T. T. "Student Handbook for General Teaching Skills Clinic." Miami: College of Education, Florida International University. 1972.

Alternative Teacher Assessment Strategies*

By

W. JAMES POPHAM

At a time when the Nation's entire educational enterprise is being subjected to increasing public scrutiny, it is not surprising that the classroom teacher—generally conceded to be the pivotal figure in most instructional settings—is being evaluated more frequently and more rigorously than ever before. Perhaps this increased stress on teacher evaluation stems from the widespread concern about accountability in education. Possibly it is the predictable concomitant of a glutted teacher market where employers can finally be choosy, when in former years their chief concern was to get an instructor, any instructor, to cover every classroom.

Whatever the cause, concerns about teacher evaluation have become far more pronounced in the past few years than at any time during this century, even though educational researchers have been continually carrying out teacher effectiveness investigations for well over 70 years. The difference in the focus of these activities provides the key to our understanding of why today's typical teacher starts to perspire a bit when someone mentions teacher competence assessment. In the old days, most teacher effectiveness researchers were searching for a suitable criterion variable which, if located, would permit them to isolate independent variables (such as teachers' personality traits, educational experience, or instructional styles) that would contribute to more effective instruction. Such investigations were accurately perceived by teachers as *research* inquiries and, as such, were not viewed as particularly threatening. Even in those instances where the attention of the investigator was clearly directed toward teacher evaluation, few teachers were very concerned. After all, even if defensible assessment techniques were discovered, it was generally held that teacher evaluation efforts would be directed toward helping teachers, never firing them. The American public had great confidence in the Nation's public schools; and, although everyone knew there were differences in the abilities of teachers, a tenured teacher's position was next to inviolate. For example, a recent search [1] of California's teacher employment records revealed that during the last 40 years not one California teacher has been dismissed on the grounds of incompetence. It is small wonder that in the past our teachers have not been too threatened by teacher evaluation activities.

But they are threatened now—and with good reason. Dissatisfied legislators in a good many states are beginning to enact penalty-laden laws which call for more stringent teacher accountability. The most celebrated of these recent teacher accountability laws is the so-called Stull Act (named after its author, Assemblyman John Stull) passed by the California legislature during the 1971 legislative session (Assembly Bill 293). The Stull Act has generated an immense amount of educational activity among California school people,[2] for its implications are serious indeed. The new law calls for the annual evaluation of all probationary teachers and the biennial evaluation of all nonprobationary teachers. The evaluations must be made on the basis, as stipulated by law, of pupil progress according to district-explicated standards of achievement in all areas of study. What has happened in California as a result of the Stull Act is that an attempt has been made to *operationalize incompetence* so that even tenured teachers can be dismissed if they are evaluated adversely. We can expect to see other state legislatures enacting comparable teacher evaluation laws in the next few years, particularly if the California experiment seems to be working.

But even if no more states establish teacher appraisal systems, there is still a strong likelihood that local districts, perhaps buffeted by school board pressures, will set up some sort of teacher evaluation sys-

* An invited working paper for a meeting of the Multi-State Consortium on Performance-Based Teacher Evaluation, New Orleans, February 25-28, 1973.

[1] Personal communication, Research Department, California Teachers Association, Burlingame, Calif.

[2] See W. J. Popham, "California's New Precedent-Setting Teacher Evaluation Law," Educational Research, Vol. 2, No. 7, July 1972, pp. 13-15.

tem. In view of these developments at the state and local levels it does not require much prescience to be able to forecast an increasing need for the technical devices and procedures required for effective teacher appraisal systems.

Although it is generally assumed by most laymen (and many legislators) that educators currently possess adequate devices for use in evaluating a teacher's instructional effectiveness, nothing could be farther from the truth. The history of teacher effectiveness research is replete with failure after failure in efforts to devise defensible indicators of how well a teacher teaches. Space limitations preclude an exhaustive analysis[3] of the limitations of previously tried assessment schemes, but each of the chief contenders, that is, ratings, observations, and pupil test performance, have fatal defects.

Ratings

Briefly, the difficulty with ratings of teacher effectiveness (characteristically supplied by administrators, but also obtainable from students, peers, etc.) is that different raters have different notions regarding what it is that constitutes good teaching. The same teacher who is rated high by one individual because of such factors as "flexible interaction with learners" and "personable, informal rapport with class" may be rated low by another individual because of "poor discipline" and "classroom anarchy." We all use our private value matrix in judging whether good teaching has taken place, and when we try to pool these disparate sets of rater expectations, chaos is the characteristic result. How many times, for example, has a classroom teacher been rated negatively by a principal because the teacher was conducting class in a way other than the manner in which the principal recalled his/her lustrous days in the classroom. Yet that same teacher may receive a positive rating by the district office supervisor who has a different idea of how teaching should be carried on. It has been observed that one person's humorist is another person's smart aleck. Similarly, one rater's Mr. Chips is another rater's Mr. Peepers.

Observations

With respect to systematic observations of the teacher's classroom behavior, we encounter an interesting assumption. It runs as follows: if certain process variables can be found to correlate positively with desired outcome variables, then by ascertaining whether those process variables are present, on that basis alone we can make judgments regarding the desired outcomes. For example, if it is discovered that a teacher's provision of practice opportunities for learners generally results in desirable learner attainment, then proponents of observational teacher evaluation schemes would contend that we can, at least in part, evaluate teachers on the basis of the degree to which they provide practice opportunities.

The trouble with the logic of this approach is its tendency to force one to the position that the process variables scrutinized by classroom observers are not only necessary for securing worthwhile results with learners, but that they are essentially sufficient. For if the phenomena observed; e.g., amount of teacher talk, are viewed as *means* to an end, why not assess the end results directly without encountering the measurement noise associated with the extra assessment step. For although no upstanding classroom observation devotee will ever assert that those behaviors observed are without exception associated with desired outcomes in learners (such as important cognitive or affective changes), the logic of the observation strategy pushes us to place greater reliance on means-end predictive relationships than the current sophistication of our observational techniques permits. If we are really interested in the ends, why not focus our assessment energy on them?

A second difficulty with observation-based approaches to teacher appraisal is that although a teacher may display optimal use of the classroom behaviors called for in the observation system, there may be deleterious factors present, factors not built into the observation structure, whose presence will essentially cancel out the positive features of the teacher's classroom behavior. The only way to head off this assessment difficulty would be to build an observation system so exhaustive that it could pick up all (or most) negative process variables, but by that time the system would be too vast to be practical.

Another difficulty with observational approaches to the assessment of teacher effectiveness is that whereas they might prove useful in identifying some classroom practices which in general will yield beneficial results with learners, the teacher evaluation game demands personal and particular decisions, not general guidelines. A particular teacher working toward particular goals with particular students in a particular setting

[3] For a more detailed examination of the strengths and weaknesses of various teacher effectiveness assessment approaches, see J. D. McNeil and W. J. Popham, "The Assessment of Teacher Competence," Chapter 7, second edition, The Handbook of Research on Teaching, R. M. W. Travers ed., Macmillan, 1973.

may break all the process guidelines and yet achieve superb results. The particularized interaction effects are too subtle for our currently unsophisticated observation systems. There have been several outstanding pro football quarterbacks whose passing form looked abysmal, yet when the receiver arrived at the appointed spot the ball was always there waiting.

Finally, there is considerable danger that when the stakes are high enough (and job security represents a big bet), many teachers will "fake good." Observation evaluation systems are particularly susceptible to such faking, for in these days of openly described criteria we can expect teachers to know what factors will be involved in the observation system. Indeed, any diligent and legally informed teachers organization should be easily able to unearth the observation dimensions involved. Having been apprised of what practices yield positive evaluations, is it so unrealistic to expect that teachers will tend toward the use of those practices when under observation? Of course, if one wished to employ constant monitoring of classroom behavior through such devices as closed circuit television, then such fakeability fears would be vitiated, but by then most schools would have been closed permanently because of the anti-1984 teachers' strikes.

Pupil Test Performance

The chief deficiency with the use of student test performance as an index of teacher proficiency has generally been that the wrong kinds of tests were employed. Since 1900 most teacher effectiveness research in which pupil test performance was employed as a criterion variable involved the use of standardized achievement tests. Since most standardized tests were designed to serve a different purpose; namely, to permit comparisons among individual learners (not among teachers) they invariably resulted in a "no significant different" outcome.

The difficulties with standardized or norm-referenced tests, particularly for teacher evaluation, have been treated elsewhere,[4] but their two most visible defects can be briefly identified. First, since commercially developed standardized tests must serve students throughout an entire nation, the generalized nature of their content coverage is often inconsistent with local curricular emphases. Incongruent measurement and curriculum results in misleading data. Second, certain psychometric properties of norm-referenced tests (such as their heavy reliance on producing among-learner variance) leads to tests which are sometimes insensitive to detecting the results of high quality instruction.

In the past few years the development of criterion-referenced (or mastery) tests offers teacher evaluators an alternative to standardized tests for assessing an instructor's impact on learners. The judicious employment of criterion-referenced tests for teacher evaluation purposes is only beginning to be seriously investigated.

Teaching Performance Tests

In the mid-sixties, the writer had reached a point of frustration regarding teacher effectiveness assessment devices and, after a reappraisal of alternative assessment strategies, had proposed the development of an alternative approach to solving this problem; namely, through the use of a teaching performance test. Two separate projects[5] were supported by the U.S. Office of Education, each designed to develop and attempt to validate teaching performance tests in different subject matter fields. While the rationale underlying the teaching performance test strategy, as well as the detailed results of these two projects are supplied elsewhere,[6] a brief description of the performance test approach can be supplied here.

One of the major difficulties in comparing teachers for purposes of instructor evaluation is that different teachers have different instructional emphases, thereby making across-the-board comparisons misleading. The teaching performance test counteracts this problem by providing an identical task for different instructors; namely, the ability to accomplish prespecified instructional objectives. The teaching performance test is built on the general premise that one chief reason for a teacher's existence in the classroom is to bring about worthwhile changes in students; that is, changes in their knowledge, attitudes, skills, etc. To the extent that this is true, then one criterion by which a teacher should be judged is his or her ability to bring about such changes. By providing identical instructional

[4] See, for example, W. J. Popham, "Domain-Referenced Measurement and Teacher Evaluation," Education Technology, in press: Robert Glaser, "A Criterion-Referenced Test," Criterion-Referenced Measurement: An Introduction, W. J. Popham, ed., Educational Technology Publications, Englewood Cliffs, N.J., 1971, pp. 41-51.

[5] Performance Tests of Instructor Competence for Trade or Technical Education, USOE Cooperative Research Contract No. OE-5-85-051; Development of a Performance Test of Teaching Proficiency, USOE Cooperative Research Contract No. 3200.

[6] W. J. Popham, "Performance Tests of Teaching Proficiency: Rationale, Development, and Validation." American Educational Research Journal, January 1971, 8 (1), pp. 105-117.

objectives for teachers, then giving the teachers an opportunity to accomplish those objectives using whatever instructional techniques they wish, a measure of the teacher's ability to accomplish given objectives can be provided. One might wish to argue that the better achiever of given objectives will also be the better achiever of his/her own objectives, but this is a question which can be answered empirically. If one simply decides that an important criterion of teaching is the ability to accomplish instructional objectives, then teaching performance tests would appear to have some utility in a data-based evaluation matrix.

The steps involved in a teaching performance test are these: (1) the teacher is provided with an explicit instructional objective (and sample test item) along with any background information necessary to become familiar with the subject matter related to that objective; (2) the teacher plans a lesson designed to accomplish the objective; (3) the teacher instructs a group of learners, typically a small group of learners for a short period of time; (4) the learners are posttested with an examination based on the objective. The examination has not previously been seen by the teacher but its nature is readily inferable from the objective (and sample test item) previously given to the teacher.

In the USOE-supported research studies described above, the purpose of developing the performance tests was primarily research-oriented; that is, it was anticipated that these devices would be employed principally for research purposes such as the identification of relevant independent variables. Consistent with that intent, the performance tests involved in those investigations consumed a fairly large amount of learner instructional time, ranging from 4 to 10 hours. At the conclusion of those investigations, it became clear that if teaching performance tests were to prove practical for teacher evaluation or instructional improvement purposes, they would have to be developed for much shorter periods of instructional time. As a consequence, the writer's recent development work with performance tests has featured instruments which take only 15 minutes of instructional time and are designed to be used with small groups of adults or younger learners. These teaching performance tests, frequently referred to as instructional mini-lessons, superficially appear comparable to the microteaching exercises developed at Stanford University some years back. In rationale, however, they are quite different. The Stanford microteaching lessons emphasize the teacher's acquisition of process skills; e.g., good questioning techniques. The instructional mini-lessons referred to here, on the other hand, focus more heavily on the results of the teaching than upon the instructional procedures themselves.

During the past few years teaching performance tests have been employed both in preservice and inservice teacher education settings.[7] Generally speaking, these performance tests have been of the short duration alluded to above; i.e., 15 to 20 minutes in length. But while these devices appear useful in instructional settings, for example, in helping prospective teachers become more facile at accomplishing prespecified instructional objectives, their utility for purposes of teacher evaluation has been largely unstudied.

In a recent paper[8] Glass has proffered the notion that teaching performance tests may have insufficient reliability to permit their effective use in teacher evaluation enterprises. Glass cited several investigations in which the reliability of teaching performance tests was clearly inadequate. Several of the investigations cited, however, had been conducted as doctoral dissertations or by novice researchers. The reliability of teaching performance tests is as yet a seriously unstudied matter. For one thing, the teaching performance tests used in these investigations have been constructed on an almost opportunistic basis, that is, whatever topics, objectives, etc., have come to the investigator's mind. No attempt has been made to carefully delineate the truly critical dimensions in teaching performance tests. Beyond that, only one investigator[9] has carefully attempted to study the reliability of even these ill-defined performance tests. Results of this investigation will be reported by Millman at the 1973 meeting of the American Educational Research Association. Examination of the Millman findings suggests that the reliability evidence, once again, is not encouraging. But, as indicated above, the nature of the performance test employed in that investigation was not rigorously explicated.

When this paper was solicited as one of several dealing with the "state of the art" in the assessment strategies suitable for performance-based teacher education, I had just completed the final draft of an American

[7] W. J. Popham, Applications of Teaching Performance Tests to Inservice and Preservice Teacher Education. A paper presented at the annual meeting of the American Educational Research Association. New Orleans. February 26-March 1, 1973.

[8] Gene V. Glass. Statistical and Measurement Problems in Implementing the Stull Act, Stanford University Invitational Conference on the Stull Act, October 1972. Palo Alto, Calif.

[9] Jason Millman, Psychometric Characteristics of Performance Tests of Teaching Effectiveness. A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 1973.

Educational Research Association (AERA) paper describing a set of minimal competencies for a performance-based teacher education program.[10]. I had even sketched alternative assessment procedures for each of the competencies. Now I just couldn't bring myself to rewrite the paper or even to subtly paraphrase my original paper. I try to restrict my paraphrasing talents to the writing of others, not my own.

Accordingly, in the present effort I have attempted to focus exclusively on the major assessment alternatives for teacher competence appraisal. Since if performance-based teacher education programs cannot demonstrate that their competency-armed products are indeed better teachers, then the performance-based teacher education folk had best fold up their competencies and slip away into the night.

[10] W. J. Popham, Identification and Assessment of Minimal Competencies for Objectives-Oriented Teacher Education Programs. A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 1973.

Conference Participants

Applications Of Teaching Performance Tests To Inservice And Preservice Teacher Education*

By

W. JAMES POPHAM

Irrespective of whether the programs are referred to as "performance-based," "competency-based," or "skill-focused," the advocacy of teacher preparation schemes designed to promote measurable capabilities of instructors is beginning to be quite fashionable among teacher educators. A recent survey [1] of California teacher education institutions found that two-thirds of the institutions participating in the survey indicated they were engaged at least partially in competency-based teacher preparation. There is even a multi-state consortium [2] on performance-based teacher education featuring such forward-looking states as Arizona, Florida, Minnesota, New York, Oregon, Texas, Utah, Vermont, and Washington. Surely, if most teacher educators are not caught up in the competency-based game, then we are at least witnessing the pregame warmups.

Whether competency-based teacher education will become something other than one of those ephemeral fads so common in teacher education remains to be seen. Although many teacher educators are quick to join the competency-based movement at the verbal level, few are willing to devote the requisite energy to devising the criterion measures without which the approach is only rhetoric. In the recent survey [3] of California teacher education colleges and universities, no institution reported satisfactory assessment procedures for a competency-based program. As with so many educational innovations, advocacy is less expensive than workable implementation procedures.

Teacher Competencies as Enabling Skills

But there is another aspect of the competency-based teacher education movement which is equally intriguing. If you sit in on almost any discussion among proponents of competency-based teacher education, you will hear them describing the teacher competencies they are trying to promote as though they were ends in themselves. As with many recent religious converts, the fervor of these teacher educators for competency identification has become so all-consuming that they lose sight of what the competencies are really supposed to accomplish.

Reducing the problem to its essentials, we can see that whatever competencies a teacher acquires must be viewed as vehicles for making that teacher more effective. More effective, in this instance, means a teacher better able to help learners. Thus, the competencies most frequently identified by performance-prone teacher educators are really en route skills which should contribute to the terminal skill of being able to help learners. For example, suppose one of the teacher competencies we are trying to promote involves the teacher's ability to view real or simulated instructional situations and identify the extent to which certain instructional tactics have been employed. We assume that the teacher who can master such a skill will subsequently be able to apply this skill in real instructional situations. Such skills should thus be viewed as precursive to one's becoming an effective teacher.

Now the point of this distinction between en route and terminal competencies is that the bulk of competencies currently viewed as the staples of performance-based teacher education are well removed from those which might legitimately be viewed as terminal. Unless en route skills are constantly verified as being actual contributors to terminal skills, then we have little assurance that defensible competencies are being promoted.

---

* A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 26-March 1, 1973.

[1] Performance-Based Teacher Education, Vol. 1, No. 5, December 1972, pp. 1, 5.

[2] Multi-State Consortium on Performance-Based Teacher Education. c/o Bureau of Teacher Education, New York State Education Department, Albany, N.Y.

[3] *Op. cit.*

Some of the competencies sought by teacher educators should be closer to the terminal proficiencies we wish teachers to display. The purpose of this paper is to explore the teacher education applications of one measurement approach designed to assess such a near-terminal competency. The measurement approach under consideration is the teaching performance test.

Teaching Performance Tests: Description and Rationale

In brief,[4] teaching performance tests work as follows: An instructor is presented with one or more explicit instructional objectives (plus a sample test item) and is directed to prepare a short lesson designed to accomplish the objective(s). If the objective deals with a topic presumed to be unfamiliar to the teacher, then relevant background information is made available. After planning the lesson, the teacher instructs a group of learners (either children or adults) for a short period of time: e.g., 15 minutes. The number of learners can be as few as a half dozen or as many as an entire class. At the conclusion of the lesson the learners are given a posttest[5] based on the objective. While not previously seen by the teacher, the nature of the test is readily inferable from the objective the teacher has been attempting to achieve. Characteristically, learners are also asked to rate how interesting the lesson was. If such an interest rating is employed, the teacher is apprised of the forthcoming rating and encouraged to design a lesson which not only accomplishes the cognitive objective, but also promotes positive learner interest ratings.

Using these two indicators, an estimate is provided of the teacher's ability to promote prespecified objectives, both cognitive (as reflected by posttest performance) and affective (as reflected by the interest ratings). Now it may be argued that an unrepresentative estimate of instructional prowess is yielded by measuring a teacher's ability to promote learner attainment of prespecified objectives for a small group of learners during a brief time period. Yet, while perhaps not as representative of the real teaching world as we might wish, there are sufficient parallels with reality that such an assessment procedure may have utility for teacher educators. In particular, since it more closely approximates a terminal teaching skill than many of the competencies currently being fostered by performance-based teacher educators, it may have advantageous instructional and evaluational dividends. The remainder of this analysis will (1) set forth three distinctive applications of teaching performance tests in teacher education operations, (2) describe actual inservice and preservice situations in which performance tests have been employed, (3) identify usage guidelines derived from these experiences, and (4) discuss certain problems which have arisen in the use of teaching performance tests.

[4] J. D. McNeil, and W. J. Popham, "The Assessment of Teacher Competence," Chapter 7, The Handbook of Research on Teaching, R. M. W. Travers, ed., Macmillan, in press.

[5] If novel subject matter is employed, no pretest is typically employed. With less esoteric topics, a pretest may be utilized to identify sufficiently naive learners.

Application One: A Focusing Mechanism

It is the writer's belief that much of the educational ineffectiveness which exists in our schools can be attributed directly to teachers' preoccupations with instructional process. Far too many teachers are caught up with concerns about devising new and exciting ways of teaching, without ever verifying what effects those procedures have on children. Some teachers pridefully announce that they strive to "teach their class differently every year," never recognizing that they may be abandoning one year an approach that was truly effective the previous year. Innovations are adulated for their own sake. For instance, open schools are currently in vogue. Ten years ago, it was nongraded schools. A decade earlier, we were praising the raptures of the core curriculum.

Not that there is anything intrinsically wrong with these new instructional approaches, for surely they possess many meritorious features. It's just that too many educators succumb to the lure of an attractive instructional process without checking the quality of its impact on learners. And that, after all, should be the reason we search for better instructional procedures.

Hence, as a method of counteracting what appears to be an almost hereditary concern about instructional process, frequent use of teaching performance tests can provide a mechanism to focus the teacher's attention on the effects of instruction. Since in a performance test situation, the quality of a teacher's efforts is predicated on results achieved with learners, both cognitive and affective, it is difficult to discount the effects of instruction on pupils. For example, if you are a prospective teacher who during a semester is obliged to teach a half dozen or more mini-lessons (as short duration teaching performance tests are sometimes called), and the first concern after your lesson is a determination of

its effects on learners. it is difficult to see how you would not soon begin to view as important a lesson's impact on pupils.

This initial application, therefore, is instructional in nature. More specifically, it is designed to foster a disposition on the part of the teacher; namely, a disposition to view as important the effects of instruction on learners.

Application Two: A Setting for Testing the Value of Instructional Tactics

Even though the foregoing application; i.e., as a mechanism for focusing teachers' attention on the consequences of instruction, may have suggested that attention to instructional procedures was somehow reprehensible, such is surely not the case. We can only secure good results with learners if we use appropriate instructional processes. The trick is to apply instructional techniques judiciously in such a way that we can either verify their efficacy (in terms of effects on learners) or at least be able to make high probability guesses that a given technique will yield desirable results with pupils.

A second application of teaching performance tests involves their use as a method of allowing teachers to test the differential effectiveness of various instructional techniques. Teachers can complete a series of mini-lessons attempting to incorporate different instructional tactics, then judge their worth in terms of the results yielded with learners. For instance, suppose a teacher taught the same mini-lessons to two different groups of comparable learners, the lesson being essentially the same except that one lesson provided much opportunity for learners to practice the skills called for in the mini-lesson's objective, while the other lesson provided no such practice. The teacher could then contrast the posttest results of both lessons and begin to reach a conclusion regarding, for certain kinds of instructional objectives, the efficacy of providing relevant practice.

Not that a tactic-present versus tactic-absent design must be employed in this second application of performance tests, for a teacher can often gain insights regarding the value of a given instructional procedure from using the procedure even without the control treatment. This is particularly true when for a particular performance test there are some normative data which, even in rough terms, yield an estimate of how well teachers typically perform on the lesson. This point will be treated in more detail later.

An important aspect of this application of teaching performance tests is that a teacher need not be the actual instructor in a mini-lesson to profit from the mini-lesson teacher's experience. A typical format for the conduct of mini-lessons, either those taught to younger learners or to a group of one's peers, is the postlesson analysis session. During this session, the teacher's instructional approach is appraised in terms of its effects on learners. Many teachers report they learn as much from watching the mini-lesson teacher's lesson, then analyzing its strengths or shortcomings, as they do from actually teaching the lesson themselves.

Frequent teacher performance tests, either for teachers as mini-lesson instructors or as an object for group analysis of other's instructional efforts, can provide the focus of a consequence-oriented preservice or inservice teachers education program.

Application Three: A Formative or Summative Program Evaluation Device

Developmental work with teaching performance tests is still at such an early state that it may be imprudent to employ them for the evaluation of individual teachers. The only exception might be for isolating instructors who are extremely weak or strong in their ability to accomplish prespecified goals. Nonetheless, as a program evaluation assessment technique, performance tests may have considerable utility. Indeed, the most important use of teaching performance tests may be as instruments to aid in the appraisal of inservice or preservice teacher education enterprises.

The argument, briefly, is that if a teacher education program sets out to promote teachers' abilities to accomplish prespecified objectives, then the program can be legitimately evaluated in terms of its ability to do so. Here's how the evaluation strategy might work. At the outset of a teacher education program; e.g., a preservice credential sequence or an extended staff development institute, a representative sample (or all) of the participating teachers could complete one of two different performance tests (e.g., test X and test Z, with half the teachers completing test X and half test Z). At the close of the program, the teachers would complete the other performance test. The prediction would be that for both tests the teachers' post-program efforts would produce markedly better results than the pre-program efforts. As a summative evaluation stratagem, such an approach could yield devastating information. What happens, for instance, if a teacher educa-

tion program discovers its teachers are no better able after intensive instruction to promote learner mastery of objectives than they were prior to instruction? Surely drastic changes in the program seem warranted. Perhaps, if such failures are recurring phenomena, the program should be terminated.

If performance tests are employed at earlier stages of the program, with a view to guiding program modifications during the course of the instruction, then formative evaluation benefits can also be derived from the use of teaching performance tests.

In review, then, we have briefly examined three possible applications of teaching performance tests in connection with either preservice or inservice teacher education efforts. There will, of course, be other uses of performance tests in teacher education. For example, a professor might use teaching performance tests as a motivating mechanism, showing novice students that more skilled teachers can outperform beginners in such instructional situations. But focusing for the moment on the three application strategies described thus far, we can examine some actual utilizations of teaching performance tests.

Actual Inservice and Preservice Applications

The Jordan Complex. One of the first reported uses of teaching performance tests to aid practicing teachers took place during 1971 in the Jordan Complex, an affiliated group of urban schools in the Los Angeles City School District. Under the leadership of LaVerne Parks, complex director, groups of teachers met weekly after regular school hours to witness each other taking turns teaching mini-lessons, then discussing the merits of the teaching approaches. Lessons were divided about evenly between those designed for young children (volunteer pupils were used as learners) and those designed for adults (teachers participating in the program took turns serving as learners).

Reactions to the program, extending over several months, were quite positive. Merle Williamson and Joyce Cooper, the class leaders, gathered anonymously supplied course evaluation data indicating that 84 percent of the participating teachers felt the class had helped them (1) understand the role of instructional objectives, (2) plan lessons for given periods of time, (3) develop alternative instructional strategies, and (4) critique lessons on the basis of posttest results. Eighty-seven percent of the participants reported that the class had helped them personalize instruction. Mrs. Parks remarked, in reviewing the experience, that "One of the most exciting outcomes from this course was that teachers began to look more critically at themselves and their peers in terms of factors contributing to successful pupil results."[6]

UCLA. As might be anticipated because of the writer's affiliation, use of teaching performance tests in the teacher education program at UCLA over the past years has been fairly extensive. All three of the applications discussed in the previous section have been employed, with attention generally given to the use of performance tests as an instructional intervention.

Typically, teaching performance tests have been used with preservice candidates, usually involving lessons taught to other members of the class. The setting for these lessons is ordinarily referred to as a mini-lesson clinic and features the customary teaching-testing-analysis model. Usually one or two mini-lessons are taught during a 2-hour clinic session. During some terms we have tried to squeeze three mini-lesson assignments into a single 2-hour lab period. On other occassions, we have required the prospective teachers to generate their own objectives and tests, then try to accomplish the objectives in a short-term lesson, as an additional exercise in promoting learner goal attainment.

The most recent procedure we have employed for using teaching performance tests is described in some detail in the appendix. Briefly, it involves the use of weekly nine-student mini-lesson clinics during which one preservice credential candidate teaches six classmates while two classmates plus a teaching assistant serve as instructional analysts. In addition, each credential candidate is obliged to teach at least three mini-lessons outside of class time to small groups of adults. Thus, in a 10-week academic quarter, prospective teachers have about two dozen opportunities to serve as mini-lesson teachers, analysts, or learners.

As our use of mini-lesson clinics has increased, there has naturally been great interest in the manner in which the teacher education students were receiving them. At the end of the fall quarter, a quarter during which mini-lessons were employed according to the scheme presented in the appendix, students were asked to supply anonymous evaluations of the course at its conclusion. Little structure in the evaluation form was presented to the students, only requests to isolate parts of the course they liked most, liked least, etc. Of the 58 students who mentioned the mini-lesson clinics, 32 were positive and 26 were negative. In view of the fact

[6] Personal communication to the writer, May 15, 1972.

that none of the mini-lesson clinic leaders possessed any prior experience with mini-lesson clinics, either as participants or supervisors, hence were probably less skillful as clinic leaders than might have been wished, these results are not distressing.

One of the hopes in setting up mini-lesson assignments so that students were obliged to teach the same mini-lesson a second time (outside of class time) was that the mini-lesson teacher would profit from the clinic critique session. Hopefully, the insights gained from the analysis of the teacher's lesson would lead to instructional improvements when the mini-lesson was retaught and a comparable posttest form was used to assess learner achievement.

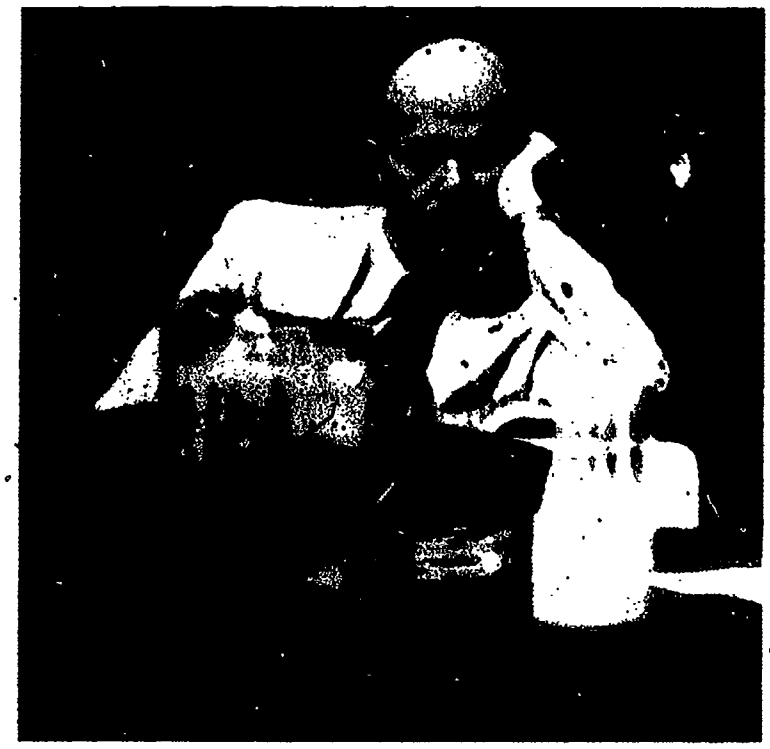
Table 1. Results of Mini-lessons Retaught After Initial Mini-lesson Clinic Analysis

	First Lesson Superior	*No Difference*	*Second Lesson Superior*
Cognitive Posttest	n=18	n=6	n=31
Interest Rating	n=13	n=9	n=33

Happily, this appears to have been the case in the fall 1972 UCLA situation for, from the reports supplied by class members (to which no grade credit was applied), both cognitive and affective performances increased more frequently on the retaught lessons than they decreased, as can be seen in table 1. Of course, these data are only suggestive in view of the fact that there were no controls exercised over the ability of learners in the retaught or initially taught lessons.

Recently we have also used performance tests in connection with program evaluation. During the fall 1972 quarter, six mini-lessons were randomly assigned to approximately a dozen credential candidates (from a particular teacher education course) at the beginning of the quarter and to a different group of credential candidates at or near the close of the quarter. In gross terms, the prediction was that the performance of these prospective teachers would be better later in the quarter, presumably after the impact of the teacher education program had worked its "beneficial" effects.

Since the mini-lessons were to be taught to classmates, care had to be taken not to involve any students in mini-lessons at the first of the quarter (either as teachers or learners) which they would encounter at the close of the quarter. Since six different mini-lessons were used, this posed no problem. Although it had been planned to have three teachers per mini-lesson on



W. James Popham

both an early-in-course and late-in-course basis, with student attrition, missed assignments, etc., a less balanced performance distribution resulted. The mini-lessons were taught to groups of approximately six learners (according to the procedural scheme described in the appendix), with the early-in-course mini-lessons occurring during the second week on the quarter. Ideally, the first week would have been used for the pretest mini-lesson, but course organizational requirements dictated that the second week was a more reasonable choice. Mini-lessons completed during weeks eight, nine, or 10 of the 10-week quarter constituted the late-in-course measures. Since all mini-lessons were completed on Fridays, this means that prior to the late course mini-lessons the credential candidates had typically experienced a minimum of about 8 weeks of instruction. The teacher education program variation under consideration involved daily 2-hour class sessions, including lectures, discussions, observations in public schools, and the mini-lesson clinic activities.

Results of the early course versus late course measurement are presented in table 2 for both the cognitive measure (posttest percentage correct) and affective measure (interest rating). All data included in table 2 are mean results for a given mini-lesson teacher, typically based on an *n* of five or six classmate-learners. It should be noted that since six different mini-lessons were employed, with different levels of difficulty and interest associated with each, interpretation of the data should focus on the columns of the table.

Table 2. Mean Results of Early-in-Course and Late-in-Course Mini-lessons Taught by Credential Candidates in a UCLA Teacher Education Program

	Mini-lessons					
	1	2	3	4	5	6
Cognitive Results	*Posttest Percentages Correct*					
Early-in-Course	72 74 80	90	36	84 90 96	82	60 74
Late-in-Course	88 90 92	73 95	36 50 65 67	86 97	82	82
	Mini-lessons					
	1	2	3	4	5	6
Affective Results	*Ratings of Lesson's Interest* (5 = hi, 1 = low)					
Early-in-Course	2.7 3.2 3.4	3.8	3.8	3.6 3.8 4.0	3.0	2.6 3.8
Late-in-Course	3.0 3.8 4.0	3.3 4.5	3.8 5.0 4.1	4.0	3.2	4.6

An examination of table 2 will reveal that the predicted results were supported by the data. For both indices; i.e., mean posttest percentages correct and interest ratings, the late-in-course performances exceeded the early-in-course performances.

By converting the results for each performance test to standard scores for that measure, with a mean of 50 and a standard deviation of 10, it was then possible to pool the data from all six performance tests and compute separate *t* tests for both the cognitive and affective results. A summary of these analyses is presented in table 3 where it can be seen that a significant difference was present in both instances, with a mean difference of .83 standard deviation unit present in the case of the cognitive measure and .97 standard deviation unit for the affective measure. Both differences, as hoped, favored the late-in-course teaching performance tests.

This represents the initial time, at least to the writer's knowledge, that teaching performance tests have been used in this manner as a teacher education program evaluation technique. It is obvious, even from examining the data table, that some refinements are in order; e.g., the disparate *n*'s in the columns, etc. Nevertheless, as an illustration of the use of performance tests for program evaluation purposes, the foregoing description may be of some utility.

California State University, Northridge. During summer 1972, two inservice workshops for approximately 50 teachers and administrators were sponsored by California State University, Northridge in coordination with Administrative Area K of the Los Angeles City Schools. Clare Rose, the instructor for both workshops, reports that teaching performance tests were used in the workshops as a vehicle for achieving one of the workshop goals; namely, that participants would be able to supervise objectives-based instructional improvement programs.

A demonstration mini-lesson was presented in each workshop, followed one week later by having all workshop participants serve either as mini-lesson teachers or students, as particular mini-lessons were taught twice by six volunteers. Prior to the second teaching of

Table 3. Analysis of Early-in-Course and Late-in-Course Mini-lesson Results, Pooled on the Basis of Standard Score Transformations

	n	$\bar{X}$	*S.D.*	$\bar{X}_{diff.}$	*t*	p*
Cognitive Results						
Early-in-Course	11	45.46	7.86	8.31	2.25	<.025
Late-in-Course	13	53.77	9.28			
Affective Results						
Early-in-Course	11	44.73	5.59	9.73	2.57	<.01
Late-in-Course	13	54.46	10.90			

* One-tailed.

each mini-lesson, an analysis of the initial lesson was carried out, Professor Rose reports that 90 percent of the mini-lesson teachers were able to promote higher posttest results on the second lesson. She indicated that 100 percent of the workshop participants reported, on anonymous end-of-session evaluation forms, that the mini-lesson activities had been valuable.

Professor Martin Levine has also employed teaching performance tests in preservice teacher education for several years at California State University, Northridge. Most recently, Professor Levine reported the following format for his use of performance tests:

> My preservice secondary education methods course meets on the campus of a participating junior high school, usually for three hours weekly (e.g., Wednesday 9-12). One or two classes of secondary pupils are assigned to the college methods class by the principal. Usually pupils in these classes are classified as "low ability learners" in need of more individual attention. Pupils are assigned at random to college trainees who are responsible for achieving prespecified instructional objectives issued by the instructor. Each trainee teaches his miniclass of three pupils for one hour during each weekly meeting of the course. Objectives deal with general study skills and thus are appropriate for both pupils who need to master this kind of objective and for college trainees who come from all of the different academic areas commonly taught in secondary schools. Trainees have a week to prepare to teach each objective. A teaching time limit of thirty minutes or less is usually set. The instructor administers a pretest and posttest. Trainees analyze their teaching effectiveness in terms of how well their pupils achieve the objectives. Trainees may re-teach objectives during subsequent meetings in cases which warrant it. From time to time, trainees combine miniclasses, with one member teaching to a prespecified objective while his peers observe the lesson for use of such instructional principles as practice and feedback. A post-observation conference is held immediately after the lesson. Usually instruction takes place in a large area such as the oral arts room or the school cafeteria where the instructor can monitor the entire process assisted by the master teachers.

User Guidelines

Based on our limited experience to date, there appear to be a few guidelines which might be of value to teacher educators considering the utilization of teaching performance tests. Some general suggestions regarding the use of performance tests are available elsewhere.[8]

First, it has become apparent that the capability of the mini-lesson analyst (when group use of performance tests is involved) is far more critical than had been anticipated. Unless the person supervising the analysis session is both convinced of the value of the activity and able to provide instructional insights (when, for example, poor learner performance occurs), then the mini-lesson may be far less profitable than possible. Too many teachers will write off mini-lesson teaching efforts, particularly if poor performances occur, as unrepresentative of their performance in a real teaching situation. Thus, one of the supervisor's missions is to clarify the parallels of the mini-lesson activity and regular classroom teaching. Further, a deft instructional analysis can provide an unsuccessful mini-lesson teacher with a promise of future success during retaught or other subsequent lessons. The mini-lesson supervisor must eliminate the frustration that follows failure if no improvement plan is presented. Skilled supervisors are so important that, in the writer's view, unless they are available in sufficient numbers (or can be trained in time) group-type mini-lesson activities should not be undertaken.

A second point relates to the use of mini-lessons with children or adults. Reports to date suggest that while a steady diet of mini-lessons for young learners is palatable, exclusive use of adult learners (e.g., teacher educator classmates or colleagues) is less acceptable. Teacher educators who, for a variety of practical reasons, may prefer to rely on mini-lessons for adult learners, should strive to provide, as a change of pace, a few mini-lessons involving younger learners.

Third, there seem to be some discernible dividends associated with providing some type of normative data, even roughly displayed, against which to interpret one's performance as a mini-lesson teacher. Without such comparative data, the teacher or supervisor is hard put to tell whether a given performance is good or bad. Referring back to table 2, it can be seen that there are clear differences, both in difficulty and probable interest, in certain of the mini-lessons. How is the mini-lesson teacher to know whether his/hers was a well designed and executed lesson if no interpretive framework is provided? We need to supply mini-lesson teachers with what any golfer needs to make the game

[8] In the teaching improvement kits distributed by Instructional Appraisal Services, Box 24821, Los Angeles, Calif. 90024.

more meaningful, a rough notice of what constitutes a par performance.[*]

Of course, normative data are not absolutely indispensable, for when a mini-lesson teacher discovers that his or her learners are all scoring around 50 percent on a task for which 90-100 percent proficiency had been anticipated, then the teacher at least has the knowledge that expectations have not been attained—and that's better than nothing. But comparative data are really a big help.

Problems

Teacher preformance tests, when employed as resources for teacher education evaluation or instruction, are in their infancy. Not surprisingly, therefore, a score of diaper-related difficulties have already arisen. And until a definitive Dr. Spock volume arrives to deal with our dilemmas, we'll have to do some trying and, unfortunately, some erring.

One of the most basic problems facing would-be users of performance tests stems from the instruments themselves. We have not yet made an acceptable effort to delineate the defining dimensions of performance tests, in terms of their content, objectives, posttest nature, background information, difficulty level, etc. Almost all of the recently developed performance tests have been devised more or less on the basis of experience and intuition. This situation needs to be rectified without delay.

A related problem is the reliability of the teacher performance tests themselves. We do not yet have sufficient data to know how many mini-lessons a teacher must attempt before we can assess the teacher's overall level of competence on such tasks. It is expected that a teacher will not rate consistently high, say, on all mini-lesson attempts because his teaching score will depend in part on the subject matter of the lesson, in part on his own teaching approach to that lesson, and in part on how well other factors have been controlled. We are optimistic that teachers on the ends of the competency continuum can be isolated for special assistance or special commendation.

[*] In the absence of clearly posted signs indicating pars (per hole) of three, four, or five, the writer's early experiences on a golf course would have suggested that an acceptable number of strokes per hole was something closer to 10, 15, or 20.

A minor problem, but one which can yield troublesome implementation difficulties, is the necessity to select topics for mini-lessons which will be viewed as important, both by the teachers who carry out the mini-lesson and by the students who are taught. In an effort to identify novel topics, thereby eliminating the need for pretesting, a few topics have been chosen which are so esoteric as to yield atypical (or all too typical) student apathy. Perhaps it may be wiser to select some main line curricular objectives and go to the trouble of pretesting prospective students in order to locate a suitably unknowledgeable learner group. Such mini-lesson topics might then be viewed as more meaningful by both teachers and learners.

Finally, the logistics problems associated with proper use of performance tests should be anticipated. An examination of the step-by-step details given in the appendix will reveal the level of organizational planning needed to head off confusion. For instance, it is highly desirable to provide teachers of unsuccessful mini-lessons with an opportunity to replan and then reteach the lesson to a different group of learners. But providing these reteaching opportunities takes a good deal of planning time. Faced with such planning frustrations, many teacher educators will be tempted to return to the less taxing, but perhaps less effective, lecture-discussion classroom format. Anticipating logistical problems can help avoid them. A competent secretary or teaching assistant can alleviate logistical distractions by working out the organizational requirements in advance.

Review

In retrospect, an attempt has been made in this paper to discuss possible applications of teaching performance tests to the activities of teacher educators. If a bias in favor of such applications was reflected in the paper, this was only natural, for such is the writer's bias. Although still a rather primitive tool, the performance test may be a valuable instrument to teacher educators. As we look at ancient man's hand axes, we may view them as incredibly simple devices, yet their impact was enormous. No strict analogy is being proposed here, only a plea to consider teaching performance tests as an additional tool in our teacher education kits. Think of how many sabre-toothed tigers we might slay.

Annotated Bibliography of Teaching Performance Test Articles

Baker, E. L., "Relationship Between Learner Achievement and Instructional Principles Stressed During Teacher Preparation," Journal of Educational Research, November 1969, 63(3):99-102. (A study is reported in which teaching performance tests are employed as a dependent measure in relationship to several independent instructional variables.)

———, Teaching Performance Tests as Dependent Measures in Instructional Research. A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 25-March 1, 1973. (Techniques for using teacher performance tests in instructional research are described along with illustrative data typifying such applications.)

Belgard, M., Rosenshine, B., and Gage, N. L., "Explorations of the teacher's effectiveness in explaining, in Bellack, A.A. and Westbury, I." (Editors), Research Into Classroom Processes: Recent Developments and Next Steps. New York, N.Y.: Teachers College Press, 1971, pp. 175-217. (Forty-three teachers taught two different 115-minute lessons to the same group of pupils. The correlation between residual class means was .47.)

Connor, A., Cross-validating Two Performance Tests of Instructional Proficiency. University of California, Los Angeles. USOE Project No. 8-I-174, 1969. (This investigation, conducted as part of a post-doctoral fellowship project by Connor, examines the correlation between the performance of 17 experienced teachers on two different performance tests. A low, positive relationship was reported.)

Glass, G. V., Statistical and Measurement Problems in Implementing the Stull Act. A presentation at the Stanford Invitational Conference on the Stull Act, October 1972. (In this paper, Glass critiques a number of previous studies regarding teaching performance tests and offers the view that their current reliability is not adequate to warrant use in the evaluation of teachers.)

Justiz, T. B., A Method of Identifying the Effective Teacher. Doctoral thesis: University of California, Los Angeles, 1969. (Univ. microfilms No. 29-3022-A.) (This investigation assessed the degree of correlation between the performances of nine teachers on two different performance tests involving different pupils. A rank order correlation of .64 was reported. A second study was also described in which five teachers taught two different performance tests to the same students. A high rank order correlation was reported for the latter analysis.)

Levine, M. G., "The Effect on Pupil Achievement of a Criterion-Referenced Instructional Model Used by Student Teachers," The Journal of Teacher Education, Vol. 23, No. 4, Winter 1972, pp. 477-481. (A 75-minute teaching performance test was used as a criterion measure to contrast the efficacy of alternative teacher education approaches.)

McNeil, J. D., "Performance Tests: Assessing Teachers of Reading." The Reading Teacher, 1972, pp. 622-627. (In this article, a teaching fair is described in which prospective elementary teachers taught different groups of three elementary school pupils in consecutive teaching situations. The investigator reported a tendency for teachers to display relatively stable proficiency on such tasks.)

McNeil, J. D., and Popham, W. J., "The Assessment of Teacher Competence," Chapter 7, The Handbook of Research on Teaching, R. M. W. Travers, Ed., Macmillan, in press. (This is a comprehensive review of research and evaluation strategies involving teacher effectiveness. Teaching performance tests are examined along with commonly employed teacher competency indicators according to various criteria for use in research and evaluation.)

Millman, J., Psychometric Characteristics of Performance Tests of Teaching Effectiveness. Paper presented to the American Educational Research Association, New Orleans, February 1973. (Describes a number of small scale investigations carried out in Pasadena, Calif., in 1971-72 in an

effort to study the psychometric properties of teaching performance tests.)

———, "Teacher Effectiveness: New Indicators for an Old Problem," Educational Horizons, Spring 1973. (A readable rationale statement endorsing the use of performance-based measures of instructor skill.)

Popham, W. J., Evaluating Instruction. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1973. (Two chapters of this book are self-instruction programs designed to provide the reader with skills needed to construct and to use teaching performance tests.)

———, "Performance Tests of Teaching Proficiency: Rationale, Development, and Validation," American Educational Research Journal, 1971, 8, 105-117. (This article describes the background of teaching performance tests and the development of such tests in the fields of social studies and vocational education during the mid-sixties. Attempts to validate the tests are reported and the validation strategy analyzed.)

———, "Teaching Performance Tests as a Vehicle for Instructor Self-Evaluation," The National Elementary Principal, Vol. 52, No. 5, February 1973. (This article describes alternative techniques of employing teaching performance tests as instruments for instructor self-appraisal.)

———, "Teaching Skill Under Scrutiny," Phi Delta Kappan, Vol. 52, No. 10, June 1971, pp. 599-602. (This article examines alternative techniques for assessing teacher skill and describes the possible role of teacher performance tests as an alternative to ratings, observations, and pupil performance on standardized tests.)

Appendix

TEAM S MINI-LESSON

CLINIC GUIDELINES

UCLA Teacher Education Laboratory
Fall 1972

Commencing with the second week of the quarter, each student enrolled in team S will be required to participate either as an instructor or a learner in a weekly mini-lesson clinic. The procedural elements of that participation will be outlined below.

One Booklet Required

During the first week of the quarter, when enrollments have settled down to some extent, a mini-lesson clinic assignment list will be distributed. Nine students will be assigned to a particular mini-lesson clinic group. Three of the students in each group will be designated as the *A* group, three as the *B* group, and three as the *C* group. (These arbitrary designations do not reflect the instructors' grading vision and will have no relationship to the final grade earned in the course.) Each student should, without delay, then go to the student bookstore and purchase one copy of the adult form Teaching Improvement Kit, which bears the same letter as the one he or she was assigned. For example, a student in mini-lesson clinic No. 13 who is designated as a member of the B group should purchase a copy of Form B of the Adult Teaching Improvement Kit. Please note that there are some optional Teaching Improvement Kits available in the bookstore for use with children, but that the required Teaching Improvement Kit, one of the three forms available (i.e., *A*, *B*, *C*), is an adult kit. With these materials in hand, the mini-lesson clinics will get underway the second week of the course in the assigned rooms.

One mini-lesson will be taught at each clinic, thus the individual assigned to teach on a given date must be present and prepared to teach. Failure to do so will result in a severe grade penalty. Yet, excused absences will occur. Therefore, all students should be ready to teach one week early in case of an unanticipated absence by the regularly assigned teacher.

General Nature of the Participation

To provide a brief overview, an individual student's responsibilities will be described. First, a student will be obliged to instruct the six students in his mini-lesson clinic group who have been assigned other letters than his/her own. For example, a C member of the mini-lesson group would be obliged to teach a lesson from the C Teaching Improvement Kit to the six A and B students in the group. Second, each student must act as a critic for two mini-lessons taught by the members of the mini-lesson group who possess the same letters. To illustrate, a B member of mini-lesson clinic group No. 7 would not only teach one B mini-lesson him/herself, but would serve (along with another B member) as a critic while two other B lessons are being taught to the six A and C students. Finally, each student will be obliged to locate one or more groups of at least three adults to whom each of the three mini-lessons in his or her teaching kit can be taught, using a form of the posttest other than that employed in the mini-lesson clinic. Thus, in summary, each student will teach one mini-lesson to team S classmates, critique two other mini-lessons being taught to team S classmates, teach three mini-lessons to adults other than team S classmates, and serve as a learner for six mini-lessons.

Procedural Specifics

Now, in more detail, here are the step-by-step procedures to be followed by each student. First, consult the mini-lesson assignment sheet to note which group, letter, and which week you have been assigned. Now read the first three chapters of the Teaching Improvement Kit (pp.1-15). You may wish to examine the mini-lesson assignment information in chapter 4. Do not examine the posttests (on green and blue sheets) which are included in the Teaching Improvement Kits. The mini-lesson clinic sessions have not been assigned to influence your grade in the course, but are specifically intended to help you improve your instructional skills. As a consequence, examination of the posttests prior to your teaching the lesson would reduce the likelihood that the mini-lesson would be beneficial to you.

On the day designated on the assignment sheet, you will be obliged to teach one of the three lessons in your improvement kit (1, 2, or 3). On that day you will instruct (for no more than 15 minutes) the six members of your group who have been assigned other letters; e.g., for the B students this would be the A and C students. The two members of your group who have been assigned the same letter will sit somewhat apart from the student group and will attempt to analyze the quality of your teaching plans, activities, etc. At the conclusion of the 15-minute lesson, the perforated copies of posttest 1 (green paper) should be removed from your Teaching Improvement Kits and distributed to your six students. You should not have seen the posttest prior to this moment. These tests should be quickly completed and scored by the mini-lesson clinic group leader. On the basis of the average scores on (1) the interest rating and (2) the posttest, discussion of the teaching should be conducted in terms of the results produced. That is, good learner performance should result in a discussion focused on instructional tactics which seemed effective; poor results should lead to a discussion focused on instructional procedures which might be altered. To assist you in judging how effective a given teaching performance was, at the rear of each Teaching Improvement Kit data are available regarding how successful other teachers have been with each mini-lesson. Results of each student's performance will be turned in, but not for grading purposes, by the mini-lesson clinic leader.

On two occasions, therefore, each student will be serving not as a learner but as a critic. On those two occasions the student (having access to the Teaching Improvement Kit under consideration) will have had an opportunity to examine the mini-lesson assignment prior to the clinic. He or she will undoubtedly have some thoughts regarding an appropriate instructional procedure. This may be beneficial during the post-lesson analysis session. It will be useful to have two people, other than the clinic leader, who are as conversant with the mini-lesson requirements as the teacher. During the analysis discussion, it is anticipated that the mini-lesson clinic leader and the two critics will take primary responsibility for isolating elements of the lesson that were particularly effective or ineffective. Remember that the appraisal of instructional means should be made chiefly in terms of learner results; that is, the averaged interest ratings and posttest scores.

The final responsibility for each team S student is to teach all three of the mini-lessons in his or her kit to another group of adults; that is, someone other than team S students. Since each of the mini-lessons in the Teaching Improvement Kits has two equivalent posttests, the team S students should not, either deliberately or inadvertently, examine posttest 2 (blue sheets). Ideally after the original teaching of that mini-lesson in the mini-lesson clinic, each student should locate a group of at least three adults (friends,

relatives, or people off the street) who would be willing to serve as students for the approximately 20 minutes involved for this assignment (15 minutes of teaching plus 5 or so minutes of testing). The mini-lesson should be taught and results summarized on the Mini-lesson Posttest 2 Summary Report Forms which have been provided (see attachment). The information called for is brief, yet should be completed in its entirety. These Mini-lesson Posttest 2 Report Forms should be turned in weekly during the 9 o'clock class meetings. In other words, an individual team S student should have an opportunity to reteach the mini-lesson he originally taught to team S students after a critique of the first lesson. Hopefully, this analysis will be useful in promoting improved learner performance. The other two lessons in the kit will, for that student, be taught for the first time. Ideally, the discussion of a classmate's teaching of that same lesson will prove useful in helping devise an effective lesson.

Other Considerations

Results of the weekly mini-lesson clinics will be made available so that students who wish to compare their performance with that of other team S members may do so. As indicated previously, there are other, optional Teaching Improvement Kits available in the bookstore for use with younger learners (elementary school age). Furthermore, mini-lessons in the adult kits can be used with mature secondary school students. Thus, it is possible to teach mini-lessons from another kit (other than the letter assigned) to high school age students. Either of these activities; that is, teaching mini-lessons to elementary or secondary school youngsters, should be considered optional for team S students. Remember, the whole purpose of the mini-lesson clinic operation is to improve the team S member's skill in accomplishing prespecified objectives with teaching procedures which are also interesting to the learners. The amount of time that you can devote to promotion of this particular competency will undoubtedly yield great benefits to your future students. As a consequence, please approach the mini-lesson clinics and outside mini-lesson assignments as a real opportunity to increase your instructional skills.

MINI-LESSON POSTTEST 2 REPORT FORM

Your name _____ Mini-lesson _____

Date mini-lesson taught _____ Today's date _____

Setting for teaching and types of learners (one or two sentences)

_____

_____

Number of students taught _____

Results: average interest rating _____

average percent correct on posttest _____

Was this the second time you taught the mini-lesson? Yes _____ No _____

If this was the second time, supply the average interest rating _____

and average percent correct _____ for the first time you taught the mini-lesson.

Comments (optional): _____

_____

_____

Below give the names and phone numbers (if available) of at least three of the students you taught.

_____

_____

Identification And Assessment Of Minimal Competencies For Objectives-Oriented Teacher Education Programs*

By

W. JAMES POPHAM

Classroom teachers have hundreds of things to do. It should follow, then, that prospective classroom teachers have hundreds of things to learn. But too many well meaning teacher educators have used these two premises to draw the conclusion that they therefore have hundreds of things they must teach, and that represents a serious error.

Given the instructional time available in typical teacher education programs, we must be more modest in our aspirations. When we ask teacher education candidates to swing an axe at every tree in the forest, they may fail to fell even a sapling. Far too many teacher education programs are predicated on a cover-the-waterfront concept: that is, give 'em the works in cultural foundations, educational psychology, and instructional methods. And the use of the verb "cover" is quite deliberate. Most teacher educators feel compelled to cover content in their courses that they perceive as germane to the teacher's responsibilities. When these professors have covered such content, they sleep easier at night. Few of these coverage-culprits ever verify whether their extensive coverage of subject matter ever results in any payoff for the teacher candidate, other than the ability to pass a memory-oriented final examination. And as in so many content-coverage courses, not just those in a teacher education sequence, what was covered one semester has faded from the student's memory by the first week of the next term.

Particularly at a time when teachers are being weighed more scrupulously on the public's accountability scales, teacher educators who persist in covering all relevent topics are probably doing an injustice to the teachers they are responsible for preparing. It makes more sense for the teacher educator to select a limited number of competencies which teachers should acquire, then focus the program's resources on making certain these skills are acquired.

Now even if this point of view (i.e., a focus on the attainment of a modest set of competencies) were assiduously followed, there would still be considerable disagreement regarding which competencies to promote or, in more general terms, what kinds of content to emphasize. Some would prefer to focus on the teacher's attainment of a wide repertoire of instructional techniques. Others might attend more directly to the teacher's becoming a more integrated human being. Still others would emphasize the teacher's acquisition of subject matter expertise. The alternative emphases are myriad.

An Objectives Orientation

The remainder of this analysis will describe a particular orientation which can be described in general terms as an outcomes-focused approach. An outcomes-focused approach emphasizes the results that a teacher's efforts produce in modifying the behaviors of learners and can be contrasted with more process-focused strategies which attend to the instructional ploys a teacher utilizes with pupils. Because instructional objectives can serve as a convenient way of describing the intended results a teacher wishes to achieve with learners, we may refer to one variant of an outcomes-focused approach to teacher education as objectives-oriented. An objectives-oriented strategy for educating teachers will be treated here.

The rationale for an objectives-oriented approach to teacher education characteristically rests on a central assumption; namely, that the *raison d'être* for a classroom teacher is to bring about worthwhile changes in learners, i.e., important kinds of improvements in their knowledge, attitudes, skills, etc. Proponents of an objectives-oriented teacher education program believe that even if a teacher lectures with consummate skill, but the students are left unchanged, the teacher has failed.

* A paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 26-March 1, 1973.

Similarly, they contend that even if the teacher has led a nondirective discussion with the artistry of Carl Rogers, but the students are basically unaffected. then the teacher has failed. The criterion, quite clearly, is not what the teacher does, but what happens to pupils as a consequence of what the teacher does. Few objectives-oriented teacher education programs are not somehow wedded to this basic view of a teacher's mission.

But how do objectives enter the picture? Well, their chief value is in helping teachers identify more clearly, prior to instruction, the kinds of changes which should be promoted in the learners. Statements of instructional objectives are nothing more than that, convenient descriptors of intended changes in learners. In the early 1960's, advocates of the oft-maligned behavioral objective endorsed such formulations vigorously because of their focus on the learner's post-instruction behavior, not on what the teacher was going to do or the content that the course would cover. It is unfortunate that some educators have become so entangled with behavioral objectives they have made them a fetish. Precise instructional objectives, in the main, are simply statements of what teachers want to happen to learners as a result of instruction. The more explicitly these intentions can be formulated, the better we can tell whether the intentions have been realized, and it is for that reason that most proponents of objectives strive for measurability as the sine qua non of an acceptable objective. But remember the central purpose of an instructional objective—it is to help an instructional planner conceptualize the kind of changes to be promoted in learners.

Proponents of instructional strategies featuring measurable objectives should forthrightly admit that their conception of the instructional process is generally one based on rational decision making. Some critics of an objectives-oriented approach denigrate such strategies as "industrial models" of education and therefore somehow unworthy of man's truly humanistic capabilities. They would prefer less systematic and intellectualized approaches, favoring instead more intuitive, dynamic models. But when Aristotle isolated the essence of man as his rational animality and held that a person's potentials were realized to the extent that those rational powers were actualized, he offered objectives-oriented teacher educators a satisfactory counter-argument. To plan one's actions on the basis of the action's likely consequences is less industrial than it is rational. To be clearheaded is not to be mechanistic. To define anticipated outcomes in advance does not relegate one to an assembly line mentality. On the contrary, to be rational in our education decisions will give our students the best chance of prospering from the education we provide them.

Are Instructional Techniques Unimportant?

With most objectives-oriented teacher education programs, it is proper to assert that a distinction is drawn between instructional means and instructional ends, with the stress typically on ends. But as anyone who has attempted to achieve a significant end will agree, it is brought about by employing appropriate means. Hence, an outcomes-focused teacher educator must be particularly attentive to instructional techniques, enhancing the teacher's skill in employing a wide repertoire of teaching tactics, for it is only through the judicious use of such procedures that significant kinds of results in learners can be attained.

Minimum Competencies

Programed instruction specialists are familiar with an approach to the development of instructional materials known as lean programing. In such a strategy the programer tries to accomplish a given instructional objective with the least possible amount of instructional stimulus material. Aside from its obvious economic advantages, lean programing carries with it a dividend when an early version program is unsuccessful. It is easier to improve a low density program by supplementing it than it is to delete segments of a high density program, for in the latter approach we may be excising the very ingredients that contributed to whatever effectiveness the program possessed.

Similarly, what is being proposed here may be described as lean competency promotion, for only three competencies of an objectives-oriented teacher education program will be recommended. Now surely teacher education candidates will learn other things as they complete their preparation programs; they may even learn some of the hundreds of things referred to at the outset of this paper. Since it will be easier to supplement a few minimal competencies than to delete from a more diverse array, it is proposed that only three such skills be emphasized in an objectives-oriented teacher education program.

The remainder of this discussion will isolate these three minimal competencies, offer some support for their importance, and describe alternative methods of assessing the degree to which each has been attained. These three competencies may be used as the guiding goals of either a preservice or inservice teacher educa-

tion effort. The differences in strategies for promoting the competencies for experienced or beginning feeders are only superficial and the differences in assessment tactics almost nonexistent.

Competency Number One

Since the chief assumption of an objectives-oriented program is that teachers should promote worthwhile changes in learners, it is not surprising that the initial competency to be fostered deals with that basic skill:

1. *Teachers must be able to achieve prespecified instructional objectives with diverse kinds of learners.*

This competency implies that a skilled teacher should, when presented with clear statements of intended changes in learners, be able to devise instructional sequences which will work; that is, which will bring about the sought-for changes in the learners. Further, the competency indicates that this skill be manifest with different kinds of learners; for example, children of differing ages, ability levels, ethnic backgrounds, socioeconomic status, etc.

The truly professional teacher not only will need to be conversant with tested instructional principles in order to design such instructional plans but will have to discover what kinds of teaching tactics personally prove effective. Not all violinists can get good music from the same fiddle. Different people must adopt different teaching styles. For some teachers a nondirective approach will work beautifully, while for other teachers such a strategy would be a disaster. It is imperative that a teacher discover what communication style, coupled with relevant instructional principles, typically results in the attainment of prespecified instructional objectives for that teacher.

Assessment Tactics

There are two prime contenders for assessing the degree to which this initial competency has been attained. The first of these involves the use of teaching performance tests (or instructional mini-lessons) whose rationale and applications are described elsewhere.[1] Briefly, a teacher is given a measurable instructional objective (typically dealing with a novel topic) along with any necessary background information needed to understand the objective. The teacher plans a short lesson (e.g., 15-20 minutes) designed (1) to accomplish the objective and (2) to be interesting to the learners. The lesson is then taught, either to a small group of six to eight learners or to an entire class. At the conclusion of the lesson, a posttest is administered to the learners. The posttest has not been previously seen by the teacher, but its nature is readily inferable from the clearly stated objective. A form requesting the learner to "rate how interesting the lesson was" is also provided. The teacher is judged on the basis of whether both the cognitive intention (the objective as measured by the posttest) and the affective intention (the promotion of learner interest as measured by the rating form) have been achieved.

The recency of serious research attention given to teaching performance tests as an evaluative tool probably renders them inappropriate at the present time for assessing individuals other than those at the extremes of a distribution; i.e., the particularly good or particularly poor goal achievers. Performance tests may also be used to evaluate the efficacy of a teacher education program by administering them on a preprogram and postprogram basis to the teachers involved. For example, suppose two teaching performance tests (X and Y) were employed. One of each could be administered to half the teachers (or teacher candidates) at the beginning and at the close of the program. The prediction would be that $\bar{X}_{pre} < \bar{X}_{post}$ and $\bar{Y}_{pre} < \bar{Y}_{post}$.

Although the reliability of different teaching performance tests has not yet been established with sufficient precision to warrant their use for individual evaluation, with more systematic delineation of the key elements constituting such tests we may find that in the future they can be used for more fine-grained analyses of individual teacher's mastery of competency number one.

A second approach to the assessment of the initial competency is to allow teachers to posit their own instructional objectives, develop a congruent mastery examination, then instruct a group of learners in order to attain the objective. Interest ratings can also be employed here. Because an objective generated by a teacher can be less readily compared with objectives pursued by other teachers, there is the additional responsibility of the educator to appraise the quality of the teacher's objective, not to mention the consonance of the test with the objective. The advantage of this

[1] W. James Popham, "Performance Tests of Teaching Proficiency: Rationale, Development, and Validation," American Educational Research Journal, January 1971, 8(1), pp. 105-117; W. James Popham, Applications of Teaching Performance Tests to Inservice and Preservice Teacher Education. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 26-March 1, 1973.

second approach is that the teachers do most of the work in generating the objectives, tests, etc. Further, because the topic need not be novel, the objective may be designed for longer periods of instructional time as part of the ongoing curriculum activities. With topics which fall within the learner's probable experience base, a pretest must be administered to establish entry behavior level.

There are, of course, a number of en route skills which a teacher should master on the way to attaining this initial competency, but by employing one, or possibly both, of these assessment tacties the teacher educator should be able to determine whether competency number one has been satisfactorily promoted.

Competency Number Two

It has been observed elsewhere that one of the consoling features of conventional instruction is that it is characteristically so impotent we need not worry too much about what its goals are. Similarly, if a teacher is not particularly proficient at accomplishing instructional objectives, then we need not be too concerned about what the teacher is attempting to do. Just suppose that a teacher has achieved competency number one, that is, has become skilled in promoting the learners' accomplishment of prespecified objectives—then it becomes extremely important to have the teacher direct this instructional prowess toward the proper goals. Accordingly, the second minimal competency of an objectives-oriented teacher education program becomes:

2. *Teachers must be able to both select and generate defensible instructional objectives.*

Since teachers who are skilled goal-achievers must become able to either generate or select worthy goals, it is fortunate that curriculum specialists are finally discarding their customary intuitive approaches in favor of more practical goal-determination procedures. For example, the current refinement of large scale educational needs assessment approaches can be translated into practical guidelines for teachers who wish to determine educational objectives in a more rational fashion. Screening of goals by the use of various taxonomies of educational objectives, such as those devised by Gagné, Mechner, Bloom, and Krathwohl, also can lead to the adoption of more appropriate goals. Without going into those technical procedures more intensively, it can be established that these are schemes now available which, albeit imperfect, can aid a teacher in the attainment of competency number two.

Since there are now available to teachers an increasing number of extant pools of instructional objectives, thereby permitting teachers to select objectives rather than be obliged to generate them personally, it seems wise to develop the teacher's proficiency in objectives selection as well as objectives generation.

Assessment Tactics

There are several procedures available for assessing the teacher's mastery of competency number two. One procedure would require teachers to generate a set of measurable objectives, then have these judged by others (using criteria of significance, suitability for learners, etc.). A description of real or fictitious learners could be given as part of the goal-generating task, and then descriptions could be examined by judges prior to the appraisal of the goals. Teachers could be asked to generate such objectives at the beginning of the teacher education program and at its conclusion. These papers could be coded and rated by judges without knowing the time at which the papers had been written. The prediction, of course, is that the end-of-course objective would be rated higher.

Another approach to assessment might consist of having a teacher select a specified number of objectives from a larger pool of such objectives, then have the selections appraised by others. As with the previous assessment approach, subsequent judgment of the teacher's objectives (either generated or selected) can be rendered according to very general or very specific criteria.

Additional assessment tactics might involve the teacher's describing, in an exam-like setting, alternative procedures for selecting or generating defensible objectives. These descriptions, as with the first two assessment tactics, might then be evaluated by judges, and if desired, on a preprogram and postprogram basis.

Variations of these approaches are possible, of course, such as having teachers themselves rate the adequacy of objectives selected by other teachers, such ratings being subsequently appraised.

Competency Number Three

The first two competencies have been highly related to instructional objectives, their determination and accomplishment. The third minimal competency of an objectives-oriented teacher education program is, un-

like the first two, quite unrelated to objectives. In fact, it is almost antithetical to a concern about objectives:

3. *Teachers must be able to detect the unanticipated effects of their instruction.*

In spite of good intentions, even combined with good intention-achieving skills, a teacher's efforts will often produce unforeseen detrimental and beneficial results with students. Hence, all of the outcomes of instruction must be considered. The teacher must be skilled in determining the totality of what happened to students, including of course what was supposed to happen.

There are several different techniques a teacher might employ to discern unanticipated effects of instruction, such as the use of (a) relatively unstructured anonymous student questionnaires (e.g., "List the best and worst things that happened to you because of this course."), (b) structured anonymous questionnaires which attempt to isolate the positive and negative effects, cognitive as well as effective, which might occur because of instruction, (c) quasi-projective techniques such as the assignment of an essay to the class dealing with topics such as "My reactions to Biology I" or "Autobiography of a U.S. History Student," and (d) the investigation of the results of a teacher's efforts by a colleague who follows Scriven's Goal Free Evaluation strategy; that is, who attempts to discover (without even knowing what the teacher's objectives were) what happened to the students.

It is imperative that objectives-oriented teacher education programs promote this third competency, for without it there is too much danger that teachers may have marvelous intentions, accomplish them beautifully, but at the same time promote harmful side effects which more than cancel out the anticipated results.

Assessment Tactics

Even more clearly than the first two competencies, this third is heavily dispositional in nature; that is, we must strengthen teachers' dispositions to attend to the unanticipated effects of instruction.

One rather primitive method of getting at this disposition is to employ an inventory such as that presented in the appendix. The rationale and scoring scheme of this inventory, "Looking at Teaching," is supplied along with the inventory. In brief, a student is asked to register various degrees of agreement with a series of statements regarding instruction, some of which deal with the use of unanticipated side effects.

On the skill side of this competency, we could always ask a teacher to describe as many ways as possible whereby a teacher who wishes to can detect such effects.

Perhaps simulation approaches offer the greatest promise with respect to ascertaining whether this competency has been mastered. Instructional situations could be presented to the teacher, either on paper, videotape, or film, in which there are clearly intended objectives plus some evidence as to the degree to which they had been achieved. In addition, there would be some subtly identified unanticipated effects of instruction. The teacher would be asked to evaluate the worth of the instruction, and a record would be made of the extent to which attention to the unanticipated side effects had been incorporated in that evaluation.

Getting teachers to describe their general evaluation strategies is another alternative, for one could then inspect such descriptions to see if, in response to this largely unstructured stimulus, unanticipated side effects were built into the teacher's analysis plan.

A Beginning

In review, an attempt was made in this analysis to defend the proposition that fewer competencies should be used as the organizing structure for teacher education programs. An objectives-oriented teacher education-approach was described and three minimal competencies for such a strategy were isolated, along with alternative assessment tactics for each.

These three competencies were identified on the basis of the writer's experience with outcomes-focused teacher education programs. They are predicated on the belief that teachers who possess such skills will be able to do a better job for the learners they attempt to serve.

The assessment tactics, however, are certainly not as sophisticated as one would wish. Hopefully, this delineation of possible assessment ploys may stimulate other objectives-oriented teacher educators to share their pet assessment devices. More importantly, perhaps, it may encourage teacher educators, both objectives-oriented and those of other persuasions, to scrutinize the adequacy of minimal skills offered by their programs and the schemes which they employ for their assessment.

Appendix

Looking at Teaching

Directions.

This inventory consists of four brief descriptions of instructional situations, each of which is followed by five statements. Please register the extent to which you agree or disagree with each of the five statements by circling the appropriate letters to the left of each statement according to the following scheme:

SA = Strongly Agree
A = Agree
U = Uncertain
D = Disagree
SD = Strongly Disagree

There are no right or wrong answers to this inventory. It represents an effort to secure your reactions to various views of instruction. Therefore, please be as candid as possible in your responses.

Situation I.

Mr. Hill is a junior high school history teacher who believes very strongly in "open education." He designs class sessions so that they are relatively unstructured, with a heavy emphasis on discussions plus individual reports of resource projects students have initiated because of their personal interests. Mr. Hill finds that students are generally responsive to his approach, but some of them register dissatisfaction that they are not learning enough to prepare them for serious high school history classes.

SA A U D SD 1. Mr. Hill has no right to emphasize open education if it deprives students of standard course coverage.

SA A U D SD 2. The type of instruction Mr. Hill is providing will generally be uninteresting to students.

SA A U D SD 3. It is impossible to combine any form of open education with adequate content coverage.

SA A U D SD 4. Mr. Hill should have devised explicit instructional plans, almost day-by-day details, prior to the beginning of the semester.

SA A U D SD 5. Mr. Hill should not try to detect any effects of his instructional scheme other than those he guessed might emerge.

Situation II.

An elementary school teacher, Mrs. Price, usually works with third- or fourth-grade children. Normally, most students who come to her class can read quite well, but 20-25 percent cannot. She devises special self-instruction learning centers for these poor readers and encourages them to go to the centers during unscheduled class time so that they can improve their reading abilities. Although the performance of these children indicates they have become somewhat better readers, they are subjected to considerable verbal abuse by the good readers in the class whenever they participate in the learning centers.

SA A U D SD 6. Self-instructional materials can be a valuable resource for any teacher.

SA A U D SD 7. Mrs. Price should have done nothing special for the poor readers coming to her class because their deficiencies were the responsibility of previous teachers.

SA A U D SD 8. Even though it was not foreseen, Mrs. Price should realize that the negative effects of the abuse they received may have been more harmful to the poor readers than whatever progress they made in reading.

SA A U D SD 9. It is normal for 20-25 percent of children to read badly; so any gains Mrs. Price can get will be all the more valuable.

SA A U D SD 10. Poor achievers must always expect to experience a certain amount of derisiveness from normal and high achievers.

Situation III.

Mr. Cohen is a high school English teacher who

plans his instruction with inordinate care. Prior to each class he details every significant level of achievement he believes students should make as a consequence of his course. He also attempts to spell out any major attitudinal or interest shifts he is attempting to promote with the pupils. At the close of the academic year he evaluates his English class totally in terms of whether these intended changes, both intellectual and attitudinal, have been produced in the learners.

SA A U D SD 11. Mr. Cohen should certainly determine whether, at the beginning of the academic year, his pupils can always display the intended behaviors.

SA A U D SD 12. If learners are informed of the clear expectations of an instructor, such as those which Mr. Cohen appears to have, they will tend to be less anxious about the learning situation.

SA A U D SD 13. Beyond the clearly delineated behavioral changes which Mr. Cohen has identified, he should discern whether there were any adverse or beneficial effects of students which he had not considered prior to instruction.

SA A U D SD 14. Mr. Cohen's careful planning, although commendable in the abstract, will probably take too much valuable energy from his actual instruction.

SA A U D SD 15. In general, humanity subject fields such as English are the least amenable to an instructional approach dependent in the prespecification of educational goals.

Situation IV.

Ms. Harold is an elementary school's music instructor who must work instructionally with children at all levels. She feels terribly overburdened with the number of youngsters she is obliged to service, thus devises extremely intensified music lessons for each grade level. Although there is little doubt that the children are learning about music, there are a number of indications that they are becoming antagonistic to music in the process. Ms. Harold behaves as though she were oblivious of these negative attitudes.

SA A U D SD 16. Ms. Harold is probably required to undertake instruction beyond what might be expected of a typical teacher, hence we should excuse any negative attitudes she might be creating.

SA A U D SD 17. The negative attitudes the children are developing are relatively unimportant, particularly because the children are learning a great deal about music.

SA A U D SD 18. Only in esthetic fields such as music and art will intensified instruction lead to student negativism.

SA A U D SD 19. Ms. Harold should abandon any emphasis on music skills and focus instead on promoting positive attitudes toward music.

SA A U D SD 20. Ms. Harold must recognize that unanticipated effects of instruction are potentially more important than intended effects and should strive to identify such effects as the negative attitudes seen here.

Scoring Directions

Looking at Teaching

This inventory is designed to detect how predisposed teachers are to detecing the unanticipated effects of instruction and use them in evaluating the quality of instruction. Of the 20 statements with which respondents are to indicate agreement or disagreement, only five deal with this question. The other 15 items are included only to camouflage the real purpose of the inventory so that respondents are not readily able to detect the socially desirable way to answer the items.

The five items and the scores associated with each response are given below. Omitted items should be given a score of 3.

KEY

Item	*Points*				
Number	*SA*	*A*	*U*	*D*	*SD*
5	1	2	3	4	5
8	5	4	3	2	1
13	5	4	3	2	1
17	1	2	3	4	5
20	5	4	3	2	1

Since a person might earn a maximum of 25 points on the basis of these five items, scores approximating 25 should be considered to reflect a predisposition to consider unanticipated side effects important in evaluating the quality of a teacher's instructional efforts.

From Commitment to Practice in Assessing the Outcome of Teaching: A Case Study[1]

By

H. D. SCHALOCK JESSE H. GARRISON BERT Y. KERSH

As experience with performance-based teacher education has accumulated, the interrelated problems of competency definition and competency assessment have come increasingly into focus. On the one hand it is recognized that teaching competence is something more than the mastery of knowledge and simple teaching skills or behaviors, but on the other it is recognized that as soon as the definition of competency extends beyond the knowledge and skill level, the matter of assessment becomes inordinately complex. In fact, in the eyes of many, it takes on properties that demand more from the technology of measurement and evaluation than that technology has at the moment to give.

As a result of this circumstance the designers of teacher education programs face a difficult choice. If a teacher education program is to be performance- or competency-based, and not be a mockery of those concepts, provision must be made for the collection of acceptable evidence of competency demonstration. Yet the technology of assessment is such that when teaching competency is defined in terms of performance in ongoing school settings, or in terms of the outcomes expected to be achieved through teaching (Turner's criterion levels 3, 2, and 1), the wherewithall to measure such outcomes simply is not available. When confronted with this circumstance most program designers adopt the simpler definitions of teaching competency and proceed with program development as if such definitions were acceptable.

This circumstance, in our opinion, helps account for the large number of teacher education programs in existence today that label themselves as being competency- or performance-based, but choose to define "competency" at the knowledge or simple skill level. Since so few institutions have moved to implement a performance-based teacher education program that is based upon higher order definitions of competency, and those that have moved in this direction have only begun to solve the problems of assessment that are associated with it, persons planning to implement such a program have no models to pattern and no access to instrumentation that will permit them to carry out such a program even if they are courageous enough to attempt it. How else can one account for the fact that only a handful of programs define teaching competency in terms of complex skill outcomes, and probably no more than half a dozen programs in the entire United States define teaching competency in terms of the ability to perform the functions of a certificated teaching position or the ability to bring about the outcomes expected from the successful execution of a teaching position?

The intricate relationship between competency definition and assessment has been recognized for a long while. In 1970, in a paper entitled "The Focus of Performance-Based Certification: Knowledge, Teaching Behavior or the Products that Derive from a Teacher's Behavior," the first author pointed to three levels of competency definition that performance-based teacher education programs could adopt (Schalock, 1970). These were referred to in general terms as knowledge outcomes, skill or "behavioral" outcomes, and product outcomes, with product outcomes referring to the outcomes to be achieved through the performance of teaching functions in ongoing school settings. Questions were also raised in that paper as to the implications of the various levels of competency definition for assessment. The authors of the elementary models made use of similar distinctions in describing alternative foci for performance-based teacher education programs (De-

[1] This paper was presented in outline form at the meetings of the Multi-State Consortium on Performance-Based Teacher Education, New Orleans, February 25-28, 1973.

Vault, et al, 1973), and Turner's refinement of this three-level distinction to one that involves six levels (Turner, 1973) has highlighted the issue even further.

Up to now, however, little progress has been made in developing assessment strategies that parallel the three- or six-level distinctions that have been made in competency definition. Turner's efforts to measure teaching skills (1965), the evolution of microteaching (Allen and Fortune, 1967), and the recent efforts of Popham in the use of what he terms performance tests of teaching proficiency (1967, 1971) provide a beginning to the kind of assessment technology that would provide for such parallelism, but taken singly or together these developments do not as yet provide what is needed to implement a performance-based teacher education program that defines teaching competency in its more complex forms.

The purpose of this paper is to describe an effort by faculty of Oregon College of Education and the Teaching Research Division of the Oregon State System of Higher Education to develop a system for assessing teaching competency that accommodates the higher order definitions of competency. The system that is being developed rests within the context of the elementary teacher education program at OCE, and takes as its point of departure the following definition: "Teaching competency is the ability to bring about the outcomes expected of an elementary teacher in a certificated teaching position."

Background

Defining teaching competency in terms of outcomes or product criteria is consistent with the specifications of the ComField model for elementary teacher education, one of the nine elementary models developed in the late 1960's under the sponsorship of the U.S. Office of Education (Schalock, et al, 1968; Schalock, et al, 1970), and with the specifications set forth in the new process standards for educational personnel development that have been adopted recently by the Oregon Board of Education ("Process Standards," 1973). Using the "Process Standards" as the document of reference, a teaching competency is defined as:

> The demonstrated ability to bring about the expected outcomes of a role or function included in a job definition;

and a competent teacher is defined as:

> One who has acquired and demonstrated the essential competencies of a professional position and integrates and utilizes them effectively in meeting the requirements of that position in accordance with its level and certification status. At each certification level, the teacher must also provide evidence that he has mastered the knowledge and skills assumed to be required for the development of his teaching competence at that level. (p. 18 of the April 12 draft of the document)

The implications of this set of definitions for the design and operation of teacher education programs have been spelled out in detail in a paper used by the Board of Education in the review process given the new standards (Schalock, 1973a).

In the fall of 1972, OCE implemented an experimental elementary teacher education program that was to serve both as a test of the feasibility of the "Oregon Process Standards," and a test of the soundness of the principles of the ComField model. It was also to serve as a context for research and development, taking as its primary objective for the first year of operation the development of an assessment system that would meet the demands of the most exacting definition of teaching competence that is possible (levels 2 and 1 in Turner's criteria), and do so within the additional constraints established by the ComField model. These include the personalization of instruction and assessment, as defined by Schalock and Garrison (1973), the systematization of program operation, as defined by DeVault (1973) and the operation of the program within the context of a teacher education consortium, as defined by the "Oregon Process Standards."

Forty-three students entered the program. Two full-time education faculty, six quarter- to full-time faculty from related subject matter areas, 43 school supervisors, and an equivalent of one full-time specialist in measurement and evaluation staffed the program. The program was limited to the measurement and evaluation staffed the program. The program was limited to the prestudent teaching aspects of professional preparation, and extended over a period of two terms (fall and winter). Students received 36 hours of college credit when they met the requirements of the program.

A relatively limited set of developmental goals were set for the assessment system during the first year of program operation. The decision was made to concentrate on the development of those aspects of the system that would permit the assessment of teaching competencies in ongoing school settings at the precertification level, moving if time permitted to the development of competency assessment procedures at the level of initial

certification.[2] As the year progressed, this turned out to mean, operationally, the development of an assessment system that functioned at two levels of competency demonstration: 1) lesson teaching; and 2) short term (2 to 5 days), full-responsibility teaching. Lesson teaching is the first and simplest context within which teaching competency is to be demonstrated in the program. Short-term, full-responsibility teaching is the next simplest context for competency demonstration, and serves as the staging context for student teaching. Short-term teaching can be engaged in only after competency has been demonstrated in lesson teaching; and student teaching can be engaged only after competency has been demonstrated in short-term teaching.

A third demonstration context received some attention during the year, but not as much as the first two contexts that have been described. This was a *student teaching equivalency demonstration context*. It required full-responsibility teaching for a 5- to 10-day period, and could be entered only under conditions of exceptional performance in short-term, full-responsibility teaching. Successful performance in the student teaching equivalency context was accepted as evidence of the level of teaching competency required to receive initial certification.

As the assessment system now stands, it represents little more than a beginning of the system that ultimately must evolve. Two major components of the system have been developed, and a third started, but all of these have undergone major revision in preparation for the second year of program operation. Undoubtedly, they will undergo at least one more major revision before they stabilize. In addition to the revision of what has already been developed, however, the system must be extended to cover the assessment of competency for purposes of initial, basic, and standard certification. This represents a major developmental undertaking for as the "Process Standards" now reads, initial certification requires competency demonstration in a 2- to 5-week full-responsibility teaching situation (student teaching); basic certification requires competency demonstration in a one- to three-term full-responsibility teaching situation (intern or protected first-year teaching); and standard certification requires competency demonstration in a 2- to 3-year full-responsibility teaching situation after the basic certificate has been received.

Finally, the system must be extended to cover the knowledges and skills assumed to be needed to perform effectively as a teacher. This includes knowledges and skills in the various subject matter areas of professional education as well as those in the subject matter areas to be taught.

It can be seen from this brief outline that the work that remains on the assessment system far exceeds the work that has been done. What has been accomplished thus far represents only the foundation of the system that will be needed in the long run to implement the kind of competency-based teacher education program that is desired at the college, or that reflects fully the specifications of the ComField model or the new "Oregon Process Standards." The work that has been done represents a beginning, however, for the basic outline of the system has been established and several of its many components have been developed and tested. We are at least on the way, and that is more than could be said a year ago. Because of this, and because so little else exists that can be drawn upon in implementing teacher education programs that incorporate higher order definitions of competency, making public what has been done thus far at OCE may be of some benefit to others. Hopefully it will be of benefit to OCE as well, for the response to what has been done may help clarify problems that are not seen or have not been anticipated or point to the work of others that could be of benefit.

Implementing a teacher education program that is genuinely representative of the ideals of a model as complex as ComField or as demanding as the "Process Standards" that have been adopted in Oregon is a monumental task, and any institution that attempts such a venture needs all the help it can get.

Before proceeding with the description of the work that has been done on the assessment system thus far, it is worth noting that the experimental program within which the assessment system is being developed was judged to be sufficiently successful in its first year of operation that OCE faculty, cooperating school supervisors and administrators, and students recommended that it be installed as the "regular" program in the elementary division of the college. This recommendation has been accepted. Operationally this means that between 275 and 300 students will move through the professional component of the program during the coming year, that approximately the same number of

[2] The recently adopted "Process Standards" for educational personnel development in Oregon call for three levels of certification: initial, basic, and standard. Competency demonstration is required at all three levels of certification. As level of certification progresses the competencies to be demonstrated increase in number and kind, and performance standards increase in difficulty.

school supervisors and 15 or so college supervisors will have to be trained to employ the assessment system that is to be used in the program, and that 20 or more subject matter specialists will have to be at least informed of the system so that they will be able to relate to it meaningfully. The implications of this decision at the level of system operation are treated in some detail elsewhere in the paper.[1]

An Initial View of System Requirements

When initially planned, the system for assessing teaching competence at OCE was designed to accommodate a number of special conditions. These included a particular definition of assessment; a particular definition of teaching competency; a commitment to the principle of gradualism in the demonstration of teaching competency; a commitment to the principle that measures coming from the system would have utility in decisions about instruction, certification, and hiring; a commitment to the principle that measures coming from the system would be of sufficient quality that a first-rate program of research could be built around them; and a commitment to develop an assessment system that could be operated within the constraints of the resources available to the college through regular funding channels. Since each of these items had major impact upon the nature of the system that evolved, each is discussed briefly in the paragraphs that follow.

The OCE Definition of Assessment

When planning the experimental teacher education program at OCE, the assessment system that was to accompany it was to be more than a measurement or evaluation system, if by measurement one means the assignment of numerals to observations and if by evaluation one means the assignment of value to numerals. It was to incorporate these two sets of operations, but include as well the concept of an information management system that serves particular decision-making functions. Assessment was seen in the context of the OCE program, therefore, as a mechanism that supports decision making. Put in other terms, it was a targeted information system. Two major classes of decisions were to be served by the system, instructional decisions and program adaptation or design (management and policy) decisions.

Given this concept of assessment, the system was to include measures of teaching competency; performance standards for competency demonstration at particular levels of certification or precertification experience; specifications as to the decisions to be served by particular measures of competency; specifications as to the structure or mechanisms to be employed in arriving at particular classes of decisions; specifications as to the form which the data were to assume to facilitate each particular class of decisions; and an information reduction, storage, and distribution/retrieval system that permits the efficient handling of the data that come from the system. The rationale for the system, and a description of its various pieces and parts, appears in a forthcoming book entitled "Exploring Competency-Based Education" (Schalock, 1973b).

The OCE Definition of Teaching Competency

The basic definition of teaching competency that guided the development of the assessment system at OCE has already been cited (see page 59). It has also been pointed out that the definition of competency adopted by OCE is consistent with the definition proposed in the ComField model for elementary teacher education, and the definition proposed in the newly adopted "Oregon Process Standards" for the accreditation of educational personnel development programs. What has not been pointed out is the host of surplus meanings that such a definition carries.

Perhaps the most troublesome of its various surplus meanings is the fact that as defined, competency is always situation specific. The performance of instructional planning and preparation functions, for example, or the performance of instructional functions or assessment functions, are always specific to a particular group of children in a particular subject matter in a particular educational setting, *and require thereby a particular set of performance standards!* The meaning of competence is also always dependent upon the complexity of the teaching task to be performed. The demonstration of teaching competence in the context of lesson teaching, for example, has a considerably different meaning than the demonstration of teaching competence in the context of short-term, full-responsibility or intern teaching. Finally, the OCE definition of teaching competence requires that competence be demonstrated

[1] A list of the products that have been developed within the OCE experimental program, including a description of the program per se, the teaching competencies pursued within the program, and the system that has been developed for the assessment of those competencies is available upon request. Those interested in obtaining this list, or any of the items referred to on it, may do so by contacting the authors.

in ongoing school settings, and that it meet designated performance standards.

While such meanings may sound complex and strange, they are not at all arbitrary, for they follow necessarily from the basic definition of competence as the ability to bring about the outcomes expected of a certificated job position. This is so because jobs are situation specific. If such meanings are hard to understand the reader is referred again to the technical paper that spells out in considerable detail the implication of a competency definition of this kind for program operation (Schalock, 1973a).

The OCE Definition of Gradualism in the Demonstration of Teaching Competency

As indicated previously, the "Oregon Process Standards" call for competency demonstration at the level of initial certification to take place in a full-responsibility teaching situation of no less than 2 weeks duration. The assumption underlying this expectation is that a demonstration context of any less demanding nature would be insufficient as a source of evidence about ability to bring about the expected outcomes of teaching. A year of full-responsibility teaching is suggested as the demonstration context for competency assessment at the level of basic certification and 2 to 3 years of full-responsibility teaching, after basic certification has been achieved, is suggested as the demonstration context for competency assessment for the standard certificate.

The successful performance of teaching functions under conditions of full-responsibility for a 2-week period is a demanding task. It is also one that is not likely to be performed successfully without prior experience in teaching. The experimental program at OCE recognized this fact, and was planned to incorporate an extensive system of precertification teaching in which teaching competency could be demonstrated within a series of graduated competency demonstration contexts. Two demonstration contexts were to be provided prior to the 2-week context in which competence was to be demonstrated for purposes of initial certification, lesson teaching, and short-term (2-5 days) full-responsibility teaching. The lesson teaching context was to require the preparation and presentation of three lessons on three separate days and in three different subject areas.

Both demonstration contexts were to be provided within what amounts to the first two terms of a three-term professional preparation sequence. Observation and informal lesson teaching were to precede the preparation and presentation of the three lessons that were to receive formal assessment. Performance standards for lesson teaching were to be based upon a summary of performance in the three lessons, and were to be met before a student was to engage in short-term, full-responsibility teaching. The ability to bring about desired learning outcomes in pupils was not to be required as a competency at the level of lesson teaching, though the assessment and display of the learning outcomes achieved in lessons were.

After competence was demonstrated in lesson teaching, a student was then to demonstrate his competence in short-term, full-responsibility teaching. Essentially the same competencies demonstrated in lesson teaching were to be demonstrated in the new context, but with more demanding performance standards. Some additional competencies were also to be demonstrated. Performance standards were to assume much the same form as they did in lesson teaching, including not being held accountable for bringing about desired learning outcomes. That was to be a requirement, however, for competency demonstration at all levels of certification.

The demonstration of competence at the level of short-term, full-responsibility teaching was to permit a student to move on to student teaching, to intern teaching, or to a student teaching equivalency examination. The basic assumption underlying the design of the graduated demonstration contexts to be included in the program was that if students were to perform effectively in a 2- to 5-week, full-responsibility teaching situation (student teaching) that provides the context for competency demonstration at the level of initial certification they would have to have a carefully tailored set of precertification teaching experiences that would prepare them to do so.

The Utility of Competency Measures For Instruction and Job Placement

From its inception the assessment system at OCE was seen as serving the purposes of instruction, certification, and job placement. Accordingly, in each step in its development the system was to be influenced by the needs of instructional staff, the certifying agency, and the personnel officers of school districts. Information obtained through the assessment system was to serve online instructional (supervision) decisions as well as certification and hiring decisions. The latter two sets of decisions were to be facilitated by the preparation of "competency profiles" that depict competencies demon-

strated in each demonstration context. To make the preparation of such profiles feasible they were to be prepared by computer.

The Utility of Competency Measures For Research

One of the major problems that has confronted research on teacher education, or for that matter research on the effectiveness of instruction, has been the lack of adequate measures on the outcome side of teaching. One of the major aims for the OCE assessment system was the development of a system of measurement that would eliminate this problem, and by so doing enable a first-rate program of research to be established at the college. In order for this to happen a long-range, systematically designed program of research was planned on the quality of the measures themselves. The results of the first year of this research, and the lines which it is to take during the coming year, are described in subsequent sections of the paper.

The Feasibility of the Assessment System, Given the Resources Regularly Available to the Teacher Education Program at OCE

While help was to be available for the development of the proposed assessment system through grants from Teacher Corps and the National Center for the Improvement of Educational Systems, U.S. Office of Education, the system was seen as having to operate within the constraints of the resources ordinarily available to the teacher education program at OCE once it was developed. Since the system obviously would require resources to operate, this meant that ways would have to be found to utilize resources presently available in different ways, or to find as yet untapped resources and enlist them in support of system application. This stance was consciously adopted for two reasons. First, OCE did not wish to develop an elaborate assessment system and then find itself in a position of being unable to operate it because it was too costly. Second, if the system was to have general utility to the field it would have to be funotional within the resources ordinarily available to most teacher education programs.

The OCE System for Assessing Teaching Competency, Year I

Starting with the view of system requirements that has just been outlined, work was begun on the actual development of the system in September 1972. Work on the system progressed simultaneously with the development of the experimental program as a whole, and as with any interdependent effort, steps taken in one arena both influenced and were influenced by steps taken in the other. Only the results of this developmental process will be described in the present paper. The process itself is being chronicled in a monograph that describes the development of the experimental program at OCE historically (Schalock, Kersh, and Garrison, in preparation).

Teaching Competencies to be Demonstrated

In order to develop a system for assessing teaching competency one has to be clear about the teaching competencies to be assessed. Fortunately, the instructional staff at OCE were reasonably clear as to the competencies that they wished to see demonstrated at the level of initial certification, so this step in the developmental process was accomplished with reasonable ease.

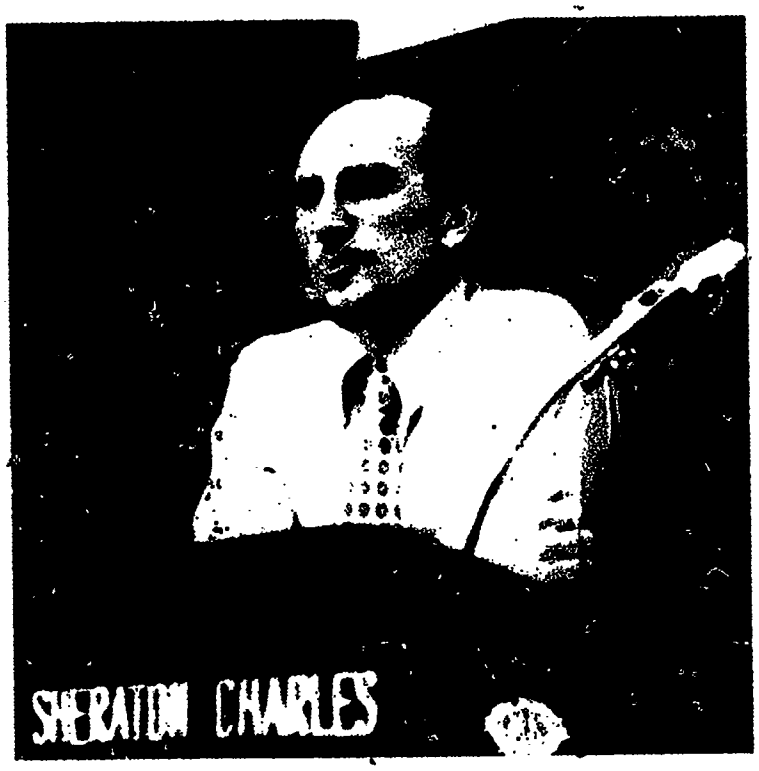
It will be recalled that the definition of competency adopted at OCE, and in Oregon generally, has to do with the ability to bring about the outcomes expected in the performance of a role or function within a certificated teaching position. As used in this definition, role or function refers to the "largest meaningful classification used in describing units of work within a teaching position" (p. 19 of the April 12 draft of the "Oregon Process Standards"). The teaching functions accepted by the OCE staff as a point of departure in developing the assessment system were as follows:

- Defining the objectives of instruction
- Adjusting instruction for the individuals involved (teacher and pupil)
- Selecting appropriate materials and procedures for instruction, given the objectives and individuals involved
- Organizing the learning environment to support instruction
- Interacting with pupils (for pupil success) in the process of instruction
- Evaluating student growth (cognitive and attitudinal)
- Defining next learning steps, and the instructional procedures that attend them, given all of the above.

These seven functions were proposed initially by Dr. Herbert Hite, now chairman of the Department of Education at Western Washington College, as content for

the ComField model (Schalock and Hale, 1968). They served as the primary organizers for instruction and assessment in the experimental program at OCE during its first year of operation.[1]

In developing the assessment system around these seven teaching functions (competencies to be demonstrated), the decision was made to obtain evidence as to how well the functions were planned as well as how well they were carried out in actual teaching. This gave rise to two reasonably distinct sets of assessment procedures, one focusing upon the assessment of lesson or curriculum plans and the other upon lesson or curriculum presentations.



H. D. Schalock

Competency Demonstration Contexts

In a program that defines competency in terms of the performance of teaching functions in an ongoing school setting the identification of the contexts in which competencies are to be demonstrated becomes as critical as the identification of the competencies themselves. In fact, it has been the OCE experience that it is not possible to think in terms of competency without thinking in terms of the context in which competency was to be demonstrated! Until these two aspects of competency definition were sorted out, little progress was able to be made in the development of the assessment system.

As indicated previously, the first year of work on the system saw the development of assessment capability in two demonstration contexts: lesson teaching and short-term (2-5 days), full-responsibility teaching. Some work was also done in relation to a student teaching equivalency demonstration context, but this was not completed. The first two of these demonstration contexts deal with precertification competency demonstration, and are designed to prepare students gradually to demonstrate the level of teaching competence required for initial certification. The third is designed as the context in which competence at the level of initial certification may be demonstrated.

Reasonably detailed specifications surround each of the competency demonstration contexts. Those surrounding lesson teaching illustrate the form and substance that they tend to take.

In partial satisfaction of requirements for completing the two-term block sequence that comprises the 1972-73 Experimental Teacher Education program at OCE, each student is to demonstrate competence in teaching elementary pupils in appropriate school settings. Each ETE student will accomplish this objective by assuming full teaching responsibility for two to five successive school days in an ongoing program of instruction, in such fashion that their performance is judged adequate by the student, the college supervisor, and the classroom supervisor in accordance with the standards set for performance in such a situation.

In preparation for the full-responsibility teaching demonstration, each student will prepare and teach three lessons, one in each of three different subject areas. The preparation and presentation of each lesson shall be done under direct supervision and in such fashion that a complete assessment can be made of strengths and weaknesses, and that the student may profit from the assessment. Satisfactory completion of all three lessons is a prerequisite to entry into the full-responsibility demonstration. The three teaching lessons are to meet the following conditions:

1. At least one lesson shall involve the teaching of reading, and at least one the teaching of art, physical education, or music;
2. At least one of the lessons shall enable the student to demonstrate ability to teach children from diverse cultural or ethnic backgrounds;
3. At least one of the lessons shall have as its primary or secondary instructional objective the learning of career awareness;
4. The desired learning outcomes for each lesson should be limited so that a lesson normally will not require more than sixty (60) minutes of instruction and not less than twenty (20);

[1] Teaching functions, of course, are always imbedded in a subject matter context, and as a consequence their assessment requires an accompanying assessment of the adequacy of the content they carry.

5. Each lesson shall be taught to a group of students numbering five or more, and at least one lesson shall involve an entire classroom or its equivalent;
6. The student will be evaluated on the basis of a specified set of teaching tasks or functions (complex skills) within each lesson, ranging from the planning of the lesson through the assessment of pupil learning from the lesson;
7. Formal lessons may be prepared and taught, at any time during the two terms allotted the experimental program, though the student needs to remember that if all requirements of the program are to be met within the two-term period, the three teaching lessons will have to be completed in time to permit the student to arrange for and complete the full-responsibility teaching demonstration;
8. An objective summary of a student's performance in the preparation and presentation of lessons will be provided to the student as soon as possible after the completion of either the planning or presentation of each lesson. The student will be permitted to offer any reaction or rebuttal to the evaluation which has been made in time for it to be of benefit to all concerned in the preparation of the next teaching lesson, or the full-responsibility demonstration; and
9. Each student will be permitted to offer any reaction or rebuttal to the final assessment made of his performance in the full-responsibility teaching demonstration before it is finally determined whether or not the student has adequately demonstrated teaching competence in terms of previously agreed upon standards.

Performance Standards for Competency Demonstration

Another aspect of the meaning of competency that had to be unraveled before progress could be made in the development of the assessment system was the matter of performance standards. This was a particularly troublesome concept for it was imbedded in both the nature of the competency to be demonstrated and the context in which it was to be demonstrated. For example, defining the objectives of instruction was a competency to be demonstrated, but there is nothing inherent in that competency descriptor that speaks to the quality expected (standard) in its performance. It also makes no reference to the context in which performance is to take place. This is equally troublesome since the performance standards for defining the objectives of instruction in the context of lesson teaching may be considerably different than in the context of short-term, full-responsibility teaching. Because of this interdependency of competency descriptor, the context in which a competency is to be demonstrated, and the performance standard set for its demonstration the task of becoming clear as to what the assessment system was to do and how it was to do it was more difficult than anticipated.

Another level of subtlety and complexity emerged in relation to performance standards as the assessment system developed. This was the distinction that had to be drawn between performance ratings and performance standards. As the system was planned initially it was anticipated that performance standards would apply to each competency that was being assessed. As the system evolved it was discovered that applying the concept of performance standards at that level of detail was simply not functional. Ratings of performance had to be applied at that level; i.e., at the level of each competency descriptor, but it turned out that performance standards seemed to apply best to performance within a particular demonstration context. Thus, as the system evolved during the first year of program operation, performance standards came to apply to performance patterns across competencies within particular demonstration contexts, rather than to individual competency demonstrations.

Such an arrangement in no way lessens the importance of individual competency assessments. These are still the central focus of the system, and their assessment provides the basis for much of the instruction that occurs in field settings. Their assessment also provides the basis for arriving at a judgment about performance standards, for these are defined in terms of individual competency assessments within a particular demonstration context. Seen in this way performance standards serve operationally as the exit requirements from a particular demonstration context, or entrance requirements to another. They also serve as the criterion measure for certification. The performance standards that were established during the first year of program operation for lesson teaching and short-term, full-responsibility teaching appear as attachment A.

The Approach Taken to Measurement

The approach taken to the measurement of individual teaching competencies was one of obtaining carefully delimited professional judgments, in the form of rating scale placements, as to the adequacy of a student's performance in a particular demonstration context. At least two separate professional judgments were obtained in relation to each competency demonstration, one from a student's college supervisor and one from

his school supervisor. An evaluative judgment was also obtained from a content specialist if a student requested it. The ratings were designed so as to accommodate the impact of setting differences on competency demonstration.

The rating scales assumed somewhat different forms for evaluating plans to evaluation presentations. In evaluating plans a three-point scale was used; in evaluating presentations, a five-point scale. This difference occurred because ratings were applied only to items in a plan that were acceptable; i.e., if an item was unacceptable it had to be modified until it was. As a consequence the ratings applied to the plan were really only the upper three scale positions of the five-point scale applied to teaching presentations.

In addition to a judgment as to individual competencies, raters were asked to provide an overall judgment as to the quality of a plan or a presentation. These latter judgments were to be made only after all individual competencies had been assessed. To illustrate the nature of the overall rating system the forms used in the application of the scales to lesson teaching are appended as attachment B.

Two other features of the rating system are worthy of note, namely, the listing of items under each competency descriptor that give focus to the descriptor and the requirement that the behavioral or product indicators relied upon by a rater in responding to an item be recorded. In the evaluation of plans, each of the focusing items was rated (which turned out to be a poor practice). In the evaluation of presentations each focusing item was attended to, but only the competency was rated. This latter strategy proved to be reasonably functional, and provided the basis for an important revision of the forms for the coming year.

The decision to ask raters to record the indicators they relied upon in making a particular judgment grew out of the need to make the ratings as objective as possible. It was felt that the inclusion of the request to record indicator statements would help bring about objectivity in two ways. First, it would force raters to at least think about, and hopefully identify, the indicators used. Second, by compiling a listing of the indicators used by different raters in different settings, a guide on indicator use could be prepared that could be of value in training raters. Unfortunately, not enough care was directed to this aspect of the rating process during the first year of program operation, and a relatively spotted record of indicator use was the result. In the coming year the rating methodology has been changed in this regard, however, and indicator usage assumes a more prominent place within it.

Data Management and Utilization

Only the most rudimentary system was developed during the first year of program operation for the management and utilization of the data coming from the competency assessment system.[*] A work-study student was responsible for the distribution of the various evaluation forms to students and college supervisors. Students were then responsible for getting the forms they received to their school supervisor, and for returning the completed school supervisor's forms to the college supervisor. A record of completed forms was maintained in each college supervisor's office for each of the students they were sponsoring, and a data summary sheet was prepared from each set of forms for use in a permanent file. The data recorded on these summary sheets were then put in computer storage for purposes of competency profile preparation and research. For illustrative purposes, the summary data form for lesson teaching appears as figure 1. The translation of this information to the computer permitted a series of methodological studies to be undertaken on the adequacy of the measures coming from the system (see the next section of the paper), and preliminary work was able to be done in the preparation of competency profiles. This task was sufficiently complex, however, that its completion was not possible during the first year of program operation.

Data management procedures that served online decision making in relation to lesson and short-term, full-responsibility teaching were left largely to the ingenuity of the college and school supervisors involved. Several guidelines to data use were provided, however. First, all conferencing with students about plans or performance in the classroom was to be data based, at least at point of departure. Second, a student had the right and obligation to quarrel with ratings or indicator statements whenever he or she felt they were inaccurate or unfair. Third, discussion of performance in the

[*] Two assessment systems were actually developed in support of the experimental program at OCE during its first year of operation. The first is the competency assessment system that is described in the present paper. The second is a program assessment system that is designed to systematically collect data on all aspects of program operation and make that data available at a time and in a form that facilitates program adaptation decisions. The program adaptation system is designed on the same principles as the competency assessment system (see pp. 8 to 14) and should therefore have the same degree of transportability.

RESEARCH SUMMARY: LESSON PREPARATION AND PRESENTATION

NOTE: The data summarized below are taken directly from LESSON PLAN and LESSON PRESENTATION evaluation forms. For some elements the data transfer is direct, that is, the element appearing on the present form has an exact counterpart on an evaluation form, and thus the rating that appears for the element here is taken directly from the evaluation form. In other cases, however, the data transfer requires manipulation, for example, further reduction or further summarization. Where this is the case, existing ratings are alway. used in the manipulations. NEW RATINGS ARE NEVER OBTAINED OR MADE FOR PURPOSES OF A RESEARCH SUMMARY.

Student's Name	Lesson Number	Person Who Provided The Rating
	* * *	

SELECTOR A: LESSON PLAN		SELECTOR B: LESSON PRESENTATION	
1. OBJECTIVES (average of the four ratings re objectives)		1. OBJECTIVES (direct transfer)	
2. ADAPTING OBJECTIVES TO LEARNER CHARACTERISTICS (direct transfer)		2. ADAPTING THE LESSON TO LEARNER CHARACTERISTICS (direct transfer)	
		3. CLASSROOM MANAGEMENT (direct transfer)	
3. SELECTING INSTRUCTIONAL MATERIALS AND PROCEDURES (average of four ratings re instruction)		4. INSTRUCTION (direct transfer)	
4. EVALUATION (average of the three ratings re evaluation)		5. EVALUATION (average of three summary ratings)	
		6. ACHIEVEMENT OF DESIRED LEARNING OUTCOMES (direct transfer)	
5. PLANNING NEXT STEPS (average of the two ratings re next steps)		7. PLANNING NEXT STEPS (direct transfer)	
6. THE FIT BETWEEN INSTRUCTOR, LESSON, AND CONTEXT (average of the two ratings re fit)		8. THE FIT BETWEEN INSTRUCTOR, LESSON, AND CONTEXT (direct transfer)	
* * * * * *		* * * * * *	
7. SUMMARY EVALUATION OF THE LESSON PLAN (average of the ratings appearing above)		9. SUMMARY EVALUATION OF THE LESSON PRESENTATION (average of the ratings appearing above)	

Figure 1. Data summary form for lesson planning and preparation.

classroom was to occur as close as possible in time to actual performance. Fourth, the college and school supervisor would confer as to the overall adequacy of the performance of a student before a student was permitted to advance to the next level of competency demonstration, even though the college supervisor was the person to make that judgment ultimately.

Within these broad guidelines all possible combinations of procedures and schedules were followed. No reduction or synthesis of the data on the evaluation forms occurred before or after discussions with students, except for the summary that was prepared for permanent file and computer use. As the reader will see in the next section of the paper the failure to direct greater attention to the management and utilization of the data coming from the assessment of lesson and short-term teaching had its consequences.

Data management procedures were more exacting in relation to the student teaching equivalency demonstration, even though much less work had been done developmentally on that demonstration context than the others. This undoubtedly reflected the fact that certification decisions were involved. The procedure that was developed to handle this level of decision required that a jury of college faculty and classroom teachers who had not served as the student's supervisors examine all data available on a student's performance and arrive at a decision concerning the match between performance and performance standards. The student's sponsor presents the data called for in this regard and provides the necessary context building that permits the jury to view the data in perspective. A segment of video tape showing the student teaching is included as a part of the student record.

This procedure was tested on five occasions during the first year of the program and found to be reasonably satisfactory. It needs to be pointed out, however, that firm performance standards were not operating on those occasions and so our experience with these procedures did not represent a full test of them.

Data on System Operation

In order to determine how well the assessment system was working in the context of the experimental program, analyses were made of the various student applications of the system, the completeness of information coming from the system, the trustworthiness of that information, etc. Studies of this nature were carried out midway through the program and then again at its completion. The data reported in the paragraphs that follow are based on these analyses.

Of the 43 students that enrolled formally in the experimental program, 42 attempted lesson teaching and 38 of these met the performance standards set for that demonstration context. All 38 of these students then attempted short-term, full-responsibility teaching, and 34 of them met the performance standards set for that context. On the basis of these figures approximately three-fourths of the students who entered the program met competency demonstration requirements for exit from it, and thereby met entry requirements for student teaching. Five of the 34 students, however, challenged student teaching through the student teaching equivalency demonstration, and three of those five were judged competent at the level of initial certification.

Taken at face value these data would suggest that the assessment system was working well. In some respects that is a fair judgment. A larger proportion of students dropped or were dropped from the experimental program than is typically the case in the nonexperimental form of the program. College and school supervisory staff also reported that the evaluations that they were able to make of student's performance were sharper and more detailed than they had ever been able to make before. Staff also reported that the data base provided by the ratings greatly facilitated instructional activities that accompanied the supervisory process.

Two other sets of data, however, force caution in interpreting how well the system worked. The first is a set of data that has to do with the conscientiousness of performance rating and documentation by various evaluators. These data are summarized in tables 1 and 2, and are informative on a number of counts. First, it is immediately clear that the rating forms were not applied in all cases, and often times when applied they were not attended to completely. Second plans are rated more consistently and more completely than presentations. Third, the conscientiousness of rating and documentation varied by class of rater, with the school supervisors generally being the most conscientious about filling out the forms completely. Finally, very few content specialists from the college faculty applied the forms to either plans or presentations, and when they did they were not overly conscientious about their use.

If taken at face value these data would suggest that the assessment system was essentially nonfunctional. This would be an over-interpretation, however, for while the application of the system obviously left much to be desired the data that appear in tables 1 and 2

Table 1. Completeness of Ratings on Competency in Lesson Planning and Presentation

Elements Rated	*College Supervisor*	*School Supervisor*	*Content Specialist*
Plans	(N = 68)	(N = 96)	(N = 20)
all elements rated			
Lesson 1	20	30	3
Lesson 2	20	23	7
Lesson 3	17	27	4
some elements rated			
Lesson 1	7	3	1
Lesson 2	1	6	2
Lesson 3	1	2	0
no elements rated			
Lesson 1	2	3	1
Lesson 2	0	1	2
Lesson 3	0	1	0
Presentations	(N = 69)	(N = 94)	(N = 20)
all elements rated			
Lesson 1	16	20	1
Lesson 2	7	19	2
Lesson 3	3	17	0
some elements rated			
Lesson 1	3	4	0
Lesson 2	1	4	1
Lesson 3	1	3	0
no elements rated			
Lesson 1	11	10	4
Lesson 2	12	8	8
Lesson 3	15	9	4

Table 2. Completeness of Ratings for Short-Term, Full-Responsibility Teaching

Elements Rated	*College Supervisor*	*School Supervisor*	*Content Specialist*
Plans	(N = 24)	(N = 29)	(N = 1)
all elements rated	18	28	1
some elements rated	4	1	
no elements rated	2		
Presentations	(N = 24)	(N = 28)	(N = 1)
all elements rated	1	16	1
some elements rated	3	9	
no elements rated	20	4	

have a number of explanations. First, the system was instigated with essentially no staff preparation. Second, some items within the system were badly in need of revision, and as a consequence many evaluators simply chose to omit them and deal only with those that made sense or were able to be managed. Third, the supervisory load on the college staff became so heavy nea. the end of the program that they were essentially unable to meet the demands that were placed upon them. This is reflected in the high proportion of ratings missing in the second and third lesson presentations, and in the short-term, full-responsibility teaching situation. Finally, no expectations were established nor held in the program for content specialists to apply the system to either plans or presentations. Students were free to ask their participation in the rating process if desired; or content specialists could ask to become involved in assessing a plan or the performance of a particular student, but this was not a matter that received a great deal of attention in the program.

In some respects, then, given the circumstances that surrounded the development and application of the system, it is possible to be delighted with the extent of the system's application and the conscientiousness with which it was applied. The data are particularly encouraging in this regard for school supervisors.

Three additional sets of data support a sense of hopefulness about the system and its operation in the context of the experimental program. The first deals with a set of analyses that were carried out to determine the sensitivity of the ratings. Two kinds of sensitivity indicators were used, the extent to which competency performance measures varied for an individual student, and the extent to which performance profiles varied across students. The assumption underlying both analyses was that in most cases variability should be found in individual competency demonstrations within the profile of any single student, and that there should be variations in competency profiles across students. It was further assumed that if such variability were observed this could be taken as evidence of the sensitivity of the measures.

The results of these analyses were in the direction desired. While some students were found to vary relatively little in the competencies demonstrated, most students varied considerably. More importantly, they tended to vary in all possible ways. For example, some students were consistently high across performance measures, some consistently low, and some both high and low. Similar variability was found between students.

The second set of data that are encouraging of the system's potential deals with the extent of agreement on ratings of student performance between independent raters. A number of analyses of this kind were made, though obviously they were limited by the incompleteness of the data as reflected in tables 1 and 2. Nevertheless, by the close of the first term of the program 22 lesson plans and 12 lesson presentations were found that were sufficiently complete to permit interrater agreements to be calculated. On the basis of these calculations, level of agreement was approximately 80 percent for the items rated in lesson plans and 75 percent for the items rated on lesson presentations. No

interrater agreements were calculated for short-term teaching.

The third set of data that are encouraging deals with rating patterns of evaluators. In these analyses, evaluator ratings across students were the basis for comparison, and were analyzed independently of the students on which the ratings were made. Pattern analyses were run that compared a) college and school supervisors ratings, b) one college evaluator's ratings with the ratings of another, c) ratings provided by school supervisors in one school with those of another, and d) ratings provided by content specialists with those provided by both college and school supervisors By and large these analyses showed that while there was some tendency on the part of all evaluators to skew the ratings toward the upper scale values, and some tendency for rating patterns to reflect the individuals doing the rating (for example, one college supervisor will tend to rate higher or with less variability than another), overall rating patterns tend to be roughly equivalent. This is especially the case as ratings being compared increase in their generality or larger numbers of ratings are compared.

This last point is illustrated by the histograms presented as figures 2 and 3. Figure 2 shows the rating patterns of two college supervisors for competence in classroom management in the context of lesson teaching. Figure 3 shows the rating patterns of all college supervisors and all school supervisors for the same measure. Even though one would expect greater similarity between the two college supervisors than between college and school supervisors, this was not the case. The greater similarity in rating patterns reflected in figure 3 can best be accounted for by the effect of large numbers entering the picture.

There tended to be less variability in rating patterns around plans on the part of all evaluators than around presentations.

As indicated throughout this discussion, the data on system performance are both encouraging and discouraging. On the encouraging side there is evidence that the measures provide reasonably sensitive discriminations in relation to pupil performance, and that raters tend to provide reasonably similar ratings when observing the same students and reasonably similar patterns of ratings when observing across students. Add to this the opinions of both college and school supervisors that the data that come from the system help them make better judgments about competence and provide better instructional help than has heretofore been possible, there is reason to be hopeful about the potential of the system. It should also be noted on the hopeful side that both college and school supervisors indicated strong support for the continued use and further development of the assessment system, and that its use is not impossibly expensive. The resources invested in the development, operation, and adaptation of the experimental program as a whole this past year are summarized in figure 4. Costs most directly attributable to the assessement function within the program are those associated with increased school supervisory time, development costs, adaptation costs, and research costs.

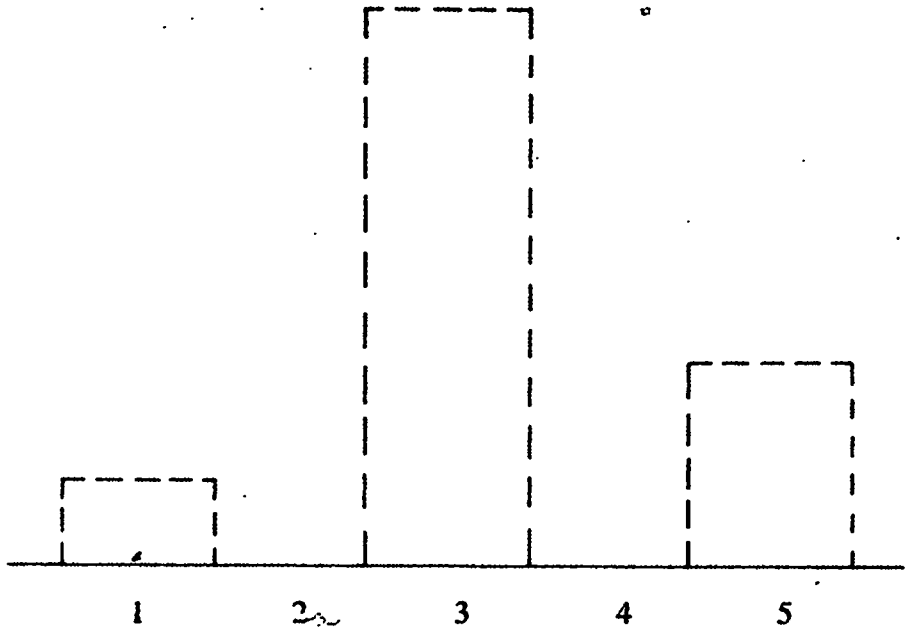
All things considered, these are reasonably encouraging data. On the discouraging side, however, there is evidence that if the system is to function effectively, and if the measures are to be of a quality that permits a great deal of confidence to be placed in them, there is still much to be done. Major revisions within the system itself must be made and an effective procedure be devised to assure care in its application. It is to the proposed modifications in the system for the coming year that we now turn.

An Expanded View of System Requirements

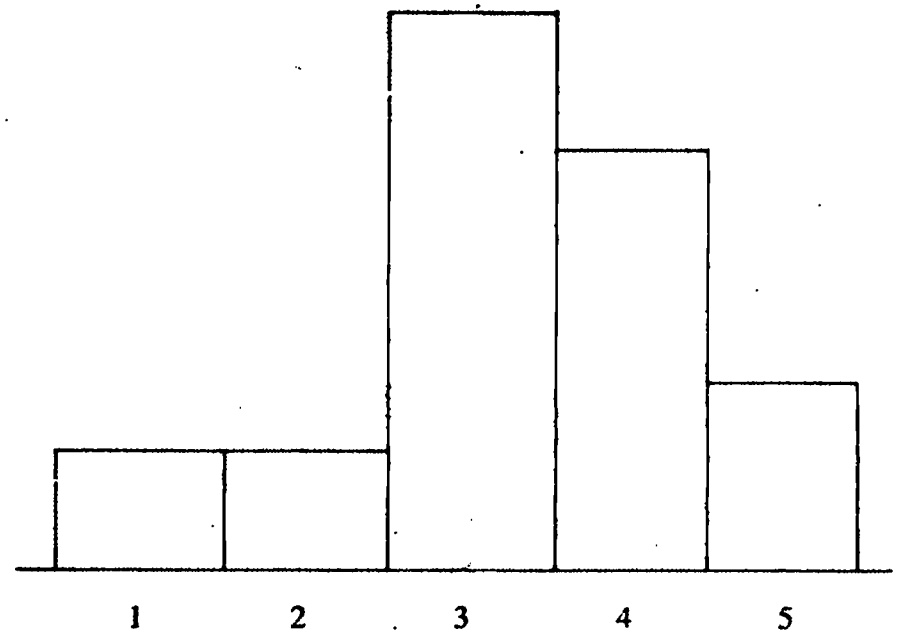
On the basis of the data just reviewed, and recommendations from college and school supervisors, students, and assessment personnel, major changes are being made in the competency assessment system for the second year of program operation. These changes will be reviewed in some detail since they represent what appears to be major advances in the methodology. To provide continuity with the description of the assessment system provided previously, the headings used in that description will be used again.

Teaching Competencies To Be Demonstrated

It will be recalled that seven teaching functions or competencies constituted the core of the experimental program during the first year of operation (see p. 64). For the coming year the list has been expanded and organized into clusters of competencies. As the list presently stands, four competency clusters are to be demonstrated in the context of lesson teaching: Planning and Preparing for Instruction; Performing Instructional Functions; Performing Assessment Functions; and Displaying Pupil Achievement. A fifth cluster of competencies has been added to these four basic clusters for demonstration in the short-term, full-responsibility teaching context and in student teaching. This is a set of competencies that has to do with Interpersonal

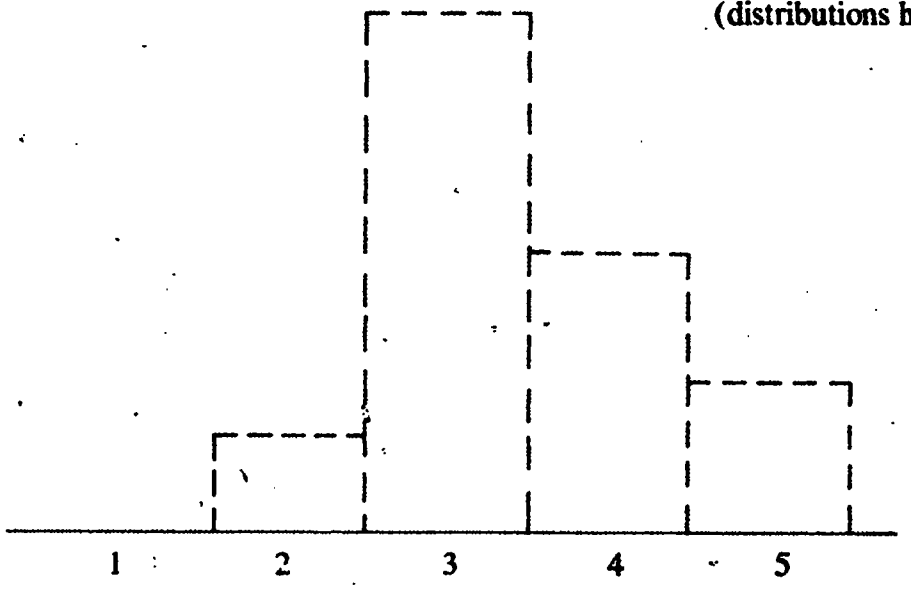


College Supervisor 01
N = 15
MEAN 1 = 3.27
STANDARD ERROR OF MEAN 1 = .26
STANDARD DEVIATION OF 1 = 1.00

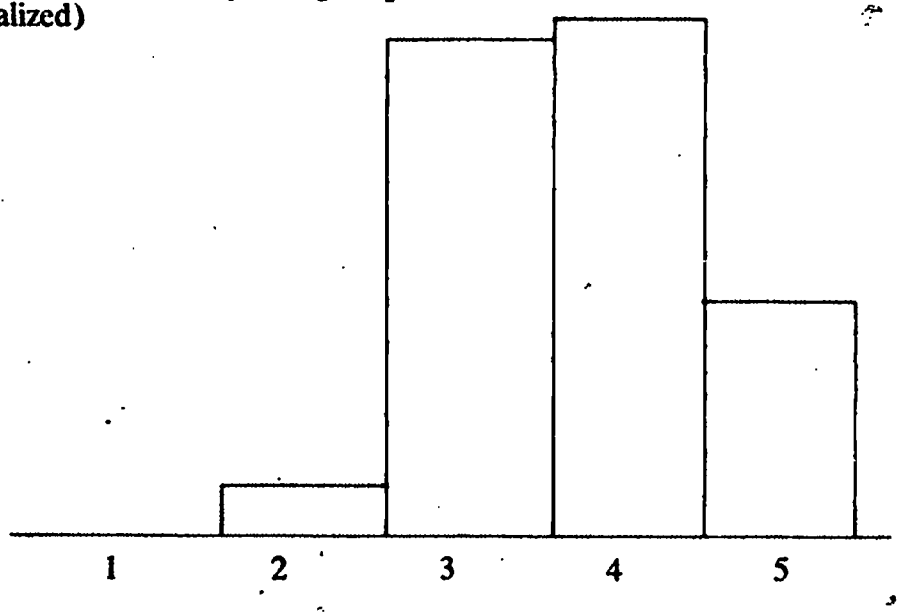


College Supervisor 02
N = 17
MEAN 1 = 3.35
STANDARD ERROR OF MEAN 1 = .23
STANDARD DEVIATION OF 1 = .97

Figure 2. The distribution of classroom management ratings on lesson presentations observed by college supervisors 01 and 02 (distributions have been normalized)



College Supervisors 01, 02
N = 32
MEAN 1 = 3.47
STANDARD ERROR OF MEAN 1 = .14
STANDARD DEVIATION OF 1 = .79



School Supervisors, All
N = 67
MEAN 1 = 3.73
STANDARD ERROR OF MEAN 1 = .10
STANDARD DEVIATION OF 1 = .78

Figure 3. The distribution of classroom management ratings on lesson presentations observed by all college and school supervisors (distributions have been normalized)

RESOURCES UTILIZED IN THE FIRST YEAR OF THE ETE PROGRAM

(44 students enrolled in the program, 37 completed it)

	Regular "Junior-Block" Program			*Experimental "Junior-Block" Program*		
	Fall	*Winter*	*Spring*	*Fall*	*Winter*	*Spring*
Program Operation						
Education Faculty	1.5 FTE	1.5 FTE		1.5 FTE	1.5 FTE	
Subject Matter School Supervisors	1.5 FTE 2 hrs. p/wk.	1.25 FTE 2 hrs. p/wk.		1.75 FTE 3 hrs. p/wk.	1.50 FTE 5.25 hrs. p/wk.	
Secretary/Clerical	Normal	Normal		+50%	+50%	
Supplies/Services	"	"		Normal	Normal	
Administration	"	"		+10%	+10%	
Program Development						
Education Faculty				.5 FTE	.5 FTE	
Assessment Faculty				1.0 FTE	1.0 FTE	
Secretary/Clerical				.75 FTE	.75 FTE	
Supplies/Services				$750	$750	
Administration				+10%	+10%	
Program Adaptation						
Education Faculty						.50 FTE
School Supervisors						6 @ 5 hrs. p/wk.
Students						12 @ 5 hrs. p/wk.
Assessment Faculty						1.0 FTE
Secretary/Clerical						1.0 FTE
Supplies/Services						$500
Program Related Research				$500	$750	$500

Figure 4. Resources utilized in the first year of the experimental program at OCE (43 students enrolled in the program, 34 completed it)

Interaction. The separate competencies to be demonstrated within these five clusters are as follows:

Competency Cluster I.
Planning and Preparing for Instruction

- Defining learning objectives and the indicators of their achievement
- Planning instructional activities, materials, and procedures
- Planning for the assessment of learning

Competency Cluster II.
Performing Instructional Functions

- Conveying the objectives of instruction
- Adapting instruction to context
- Managing the instructional process
- Managing unexpected events

Competency Cluster III.
Performing Assessment Functions

- Assessing learning before instruction
- Assessing learning during instruction
- Assessing learning after instruction

Competency Cluster IV.
Displaying Pupil Achievement

- Displaying prelesson and postlesson achievement
- Displaying learning gains that result from instruction

Competency Cluster V.
Enhancing Interpersonal Relationships

- Acting responsibly in terms of the feelings, needs, and wishes of others
- Working constructively in task-oriented situations with others

It is expected that this list of competencies will continue to expand or be refined as the assessment system is applied and tested in student teaching and intern contexts, or in contexts designed to give evidence of competency demonstration for purposes of standard certification.

Competency Demonstration Contexts

No changes are planned for the demonstration contexts that were established in the first year of program operation. In fact, with the formal acceptance of the "Process Standards" for the preparation of education personnel by the Oregon Board of Education, the demonstration contexts thus far established in the program appear to be more appropriate than initially anticipated. The competency demonstration contexts that are now proposed for the program, and their relationship to level of certification, is as follows:

Contexts for Competency Demonstration Prior to Certification

- Lesson teaching
- Short-term (2 to 5 days), full-responsibility teaching

Contexts for Competency Demonstration For Purposes of Certification

- Initial certification: student teaching (2 to 5 weeks), or student teaching equivalency demonstration (5 to 10 days)
- Basic certification: intern teaching (10-30 weeks) or protected first-year teaching
- Standard certification: 2 to 3 years of full-time teaching after the basic certificate has been received.

Performance Standards for Competency Demonstration

It will be recalled that performance standards were established during the first year of the program for competency demonstration in lesson teaching and in short-term, full-responsibility teaching only. It will also be recalled that these standards received very little formal testing. As a consequence, with one important change, the second year of the program is being entered with the same performance standards for lesson and short-term teaching that were set in the first year of the program. The change is to require a summative competency assessment for exit from the precertification phase of the program. The assessement procedure proposed is much like that outlined for assessment at the level of initial certification (exit from student teaching), for it is also to involve a jury of independent judges. These judges are to consist of college and school faculty that did not carry supervisory responsibility for a student and they are to decide whether demonstrated performance in the short-term, full-responsibility teaching situation meets the performance standards that have been set for it. A segment of video tape showing the student teaching is also to be included as part of the student's record.

One of three decisions will be reached as a consequence of this summary assessment:

- the student has passed the teaching competency requirements at the precertification level, and may enroll in student teaching or internship
- the student is eligible for student teaching or internship, but may enroll only after completing specified tasks or submitting additional evidence of teaching competence
- the student is not eligible for student teaching or internship, and must remain in the prestudent teaching part of the program until evidence can be provided that all teaching competency requirements for entry into student or intern teaching have been met.

Specific criteria have not as yet been established as a basis for any one of these decisions.

It is anticipated that with use, the performance standards that have been set for lesson teaching and short-term, full-responsibility teaching will be sharpened or in some other way modified. It is also expected that the jury system that is being proposed for summary judgment of competence in short-term teaching will be applied to all certification judgments within the program. Finally, it is expected that all certification decisions will rest heavily on evidence as to the ability of a prospective teacher to bring about the learning outcomes expected for pupils. Almost no thinking has been done within the program about the nature of competency demonstration requirements for purposes of certification beyond those that have been mentioned.

The Approach Taken to Measurement

Major refinements have been made in the approach is still one that involves ratings. Seven major changes have been made in the rating system:

- all ratings are made in terms of a five-point scale (during the first year teaching plans were rated on a three-point scale while teaching performance was rated on a five-point scale)
- ratings are provided only for competencies and competency clusters (ratings were required during

the first year of the program for the items that elaborated a competency as well)

- all ratings are to be recorded by entering a numerical score in a box opposite the competency or competency cluster to be rated, rather than marking a position on a continuous scale (it is anticipated that this procedure will force more care or attention to be given each individual judgment and its recording)
- the rating scale positions are more carefully anchored in the attributes that define each scale position
- examples of the indicators that can be relied upon in arriving at a particular rating scale judgment have been provided on the rating form
- the rating forms have been revised so that they invite easier recording of the indicators used in arriving at a particular judgment
- all competency statements, and the items that are intended to define them, have been edited and field tested for their clarity and meaning.

In combination these changes are designed to make the rating process both more manageable and discriminating.

The forms as a whole have not as yet been field tested, but initial reactions to them by college and school supervisory staff have been most encouraging. To illustrate the content and format of the new assessment forms, sample pages from the assessment battery for lesson teaching are appended as attachment C. Complete copies of the forms can be obtained upon writing the authors.

One further change has been made in the assessment forms that is of major consequence. This is their organization by the source of indicators relied upon in making judgments about the competencies being rated. Accordingly, Planning and Preparation Functions, and Achievement Display Functions, are judged in terms of *products* of a teacher's behavior; Performance of Assessment Functions are judged on the basis of a teacher's behavior per se; and Performance of Instructional Functions and the Enhancing of Interpersonal Relations are judged both on the basis of teacher behavior and pupil behavior. These distinctions will be noted in both the directions given to evaluators and in the sample indicators.

Data Management and Utilization

As yet (July 1973) the specific data management and utilization procedures to be implemented during the second year of the program have not been established, though a set of quality assurance procedures have been agreed to and an extensive program of research on the quality of the measures coming from the assessment system is being prepared. Specific procedures for the distribution and collection of forms, for the utilization of the information contained on the forms for instructional purposes, and for the utilization of that information for decision purposes relative to movement from one demonstration context to the next, are still to be defined.

Taking steps to insure the quality of the measures obtained through the assessment system is particularly critical in the coming year because of expanded use. Fifteen college supervisors, up to three hundred students and school supervisors, and a dozen or so content specialists will be applying the system throughout a half dozen school districts. Well-defined quality assurance procedures must be implemented if the data coming from the various users of the system are to be at all trustworthy.

Two strategies make up the quality assurance plan. First, it calls for a careful inservice education program to be provided on system usage. Second, it calls for systematic checks on the quality of ratings being made. These checks will be made midway through each term that the program is offered, and at the end of each term. Inservice programs will be designed on the basis of the information obtained through these checks, and data management procedures will be elaborated as needed. The research that is planned on the quality of the measures follows the same general lines as the research pursued in the first year of the program, though it will be extended in quality and scope. One addition will be the systematic study of indicator usage. The computer programs needed to carry out such research have been developed and tested so the results of these studies will be able to be acted upon as the program progresses.

The OCE Assessment System in Perspective

To those who have managed to work their way through the paper it must be abundantly clear that the assessment system being developed at OCE is a long way from completion. The parts of the system that have been developed will obviously undergo further refinement and the more complex parts of the system are yet to be developed. Problems of behavior and product sampling within demonstration contexts, performance standards for more demanding demonstration contexts, and the development of measures of competence that are trustworthy are still to be confronted. As planned

now, the completion of the total system in a form that will permit its use with known confidence is targeted for the close of the 1975-76 academic year.

While much remains to be done on the system, and major problems are yet to be resolved, a good deal of progress has been made. The basic outline of the system is complete; the basic constructs, dimensions, and methodologies of the system have been defined and implemented; data management and utilization strategies, though primitive, have been established; quality assurance mechanisms have been developed and tested; and the system in its broad outline has been found to be both acceptable and useful to college faculties, school supervisors, and students. The system has also been found to be manageable in terms of cost, particularly when developmental costs are differentiated from system operations costs. While there is obviously much that remains to be done, the rudiments of the system have at least been developed and tested. In the judgment of the authors this in itself represents a reasonable gain for the world of education.

The assessment effort at OCE needs to be viewed within still another context, however, and that is the context of research on education and teacher education. One of the great handicaps of research on the effectiveness of teachers has been the lack of strong measures of effectiveness. Without such measures, no matter how good the design or elaborate the analysis, significant relationships are not likely to be found. A weak dependent or criterion measure will defeat a strong research design and analysis everytime.

In the authors' judgment the work that has been initiated at OCE in the area of competency assessment represents a major step toward the resolution of the criterion problem in teacher effectiveness research. If all goes according to plan, 3 years from now the work begun last year will be completed, and for the first time a measurement system may be available that will meet the demands of research that can make a difference. When that time comes there can be a hopefulness about educational research that has been missing for a long while.

References

Allen, D. W. and Fortune, J. C. An analysis of microteaching: a new procedure in teacher education. In "Micro-Teaching: A Description." Stanford, Calif.: School of Education, Stanford University, 1967. Section III, pp. 1-11.

DeVault, M. V. Systems approach applications in designing teacher education programs. In Cooper, et al. (Ed.) "A Systems Approach to Program Design: Vol. II." San Francisco: McCutchan Publishing Corporation, 1973(a). pp. 19-32.

DeVault, M. V., et al., "Competency-Based Teacher Education: Problems and Prospects for the Decades Ahead." Vol. 1. Berkeley, California: McCutchan Publishing Corporation, 1973(b).

Popham, W. J. The performance test: A new approach to the assessment of teaching proficiency. "Journal of Teacher Education," 1968, pp. 19, 216-222.

Popham, W. J. Performance tests of teaching proficiency: rationale, development, and validation. "American Educational Research Journal," 1971, pp. 8, 105-117.

"Process Standards for Educational Personnel Development Programs." The Oregon Board of Education, April 12 draft copy, 1973.

Schalock, H. D. "The Focus of Performance-Based Certification: Knowledge, Teaching Behavior, or the Products that Derive from a Teacher's Behavior." Procedures of the Conference on Performance-Based Certification, Florida State Department of Education, May 1970, Miami Beach. Mimeographed. Also available in the ERIC summary of the conference.

Schalock, H. D. Implications of the Oregon Board of Education's Proposed Process Standards for the Design and Operation of Teacher Education Programs." Interpretive Paper No. 1: Defining and Assessing Teacher Competence. April 1973 (a). Mimeographed

Schalock, H. D. Notes on the concept of assessment in the context of performance-based teacher education. In R. Houston (Ed.) "Exploring Competency-Based Education." (In preparation), 1973(b).

Schalock, H. D. and Garrison, J. H. The personalization of teacher education programs. In DeVault, et al. (ed.) "New Directions in Teacher Education: Problems and Prospects for the Decades Ahead." San Francisco: McCutchan Publishing Co., 1973. pp. 33-34.

Schalock, H. D. and Hale, J. R. "A Competency-Based, Field-Centered Systems Approach to Elementary Teacher Education. Volume I: Overview and Specifications. U.S. Office of Education, Bureau of Research, Project No. 89022. Portland, Oregon: The Northwest Regional Educational Research Laboratory, 1968.

Schalock, H. D., Kersh, B. Y., and Horyna, L. L. "A Plan for Managing the Development, Implementation, and Operation of a Model Elementary Teacher Education Program." Vols. I & II. Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1970.

Schalock, H. D., Kersh, B. Y, and Garrison, J. H. "Teacher Education in Transition: A History of CBTE at Oregon College of Education. Vol. I: The Experimental Program." In preparation.

Turner, R. L. "The Acquisition of Teaching Skills in Elementary School Settings." Bulletin of the School of Education, Indiana University, 1965.

Turner, R. L. Rationale for competency-based teacher education and certification: levels of criteria. In B. Rosner (ed.) "The Power of Competency-Based Teacher Education." Prepared by the Committee on National Program Priorities in Teacher Education. New York: Allyn and Bacon, 1973. pp. 3-8.

Attachment A. Performance Standards

PERFORMANCE STANDARDS FOR LESSON TEACHING

The first and simplest context within which teaching competency is to be demonstrated in the ETE program at OCE is lesson teaching. Competency at this level of teaching must be demonstrated before a student is free to engage in full-responsibility teaching. At least three lessons must be taught for purposes of formal evaluation, and standards must be met for both their preparation and their presentation.

Standards for Lesson Planning

In preparing for lesson teaching a reasonably detailed lesson plan must be prepared and reviewed by both college and school supervisors. The standards set for the preparation of plans are *item specific* standards, that is, both the college and school supervisor must indicate, independently, that every item to be attended to in the plan has been dealt with satisfactorily. This standard must be met before the lesson can be presented to children. If a plan does not meet this standard upon its initial review, it must be revised until it does.

Standards for Lesson Presentation

The standards set for performance in the presentation of lessons are *pattern* standards, that is, they apply to the pattern of performance demonstrated in the presentation of three or more lessons. Two standards are to be applied to the performance record of a student on the three or more lessons presented:

—evidence of favorable performance on each of the teaching functions assessed in at least one of the three lessons presented;

—evidence of favorable performance on the preponderance of teaching functions assessed in the three lessons presented. Preponderance is defined here to mean at least 75 percent of the functions assessed in the course of the three lessons presented will reflect evidence of favorable performance, and no more than 25 percent of the functions assessed will reflect evidence of unfavorable performance.

PERFORMANCE STANDARDS FOR SHORT-TERM (2-5 DAYS) FULL-RESPONSIBILITY TEACHING

When students in the ETE program meet performance requirements in lesson teaching they are free to enter the first full-responsibility teaching experience that is provided in the program. This is what is termed a short-term, full-responsibility teaching context. This experience requires a student in the program to assume full responsibility for planning and carrying out instruction in the schools for a minimum of 2 days and a maximum of 5.

Three kinds of standards are applied to a student's performance in short-term, full-responsibility teaching. Two of these correspond to the standards applied to lesson teaching. The third set of standards pertains to the utilization and management of affect.

Standards for Curriculum Planning

As in the case of individual lesson teaching, a curriculum plan for short-term, full-responsibility teaching must be approved before teaching can be undertaken. This requires that a reasonably detailed curriculum plan be prepared for a 2- to 5-day demonstration (as used here a curriculum plan consists of a number of individual lesson plans, and the relationships if any between them), the curriculum plan be reviewed by both college and school supervisors by lesson plan, and that both the college and school supervisors must indicate, independently, that each of the items to be attended to in the plan as a whole has been dealt with satisfactorily. In keeping with the generally more demanding requirements of the short-term, full-responsibility demonstration all lessons to be presented within the 2- to 5-day teaching period must meet acceptable standards before teaching can begin.

Standards for Curriculum Presentation

Standards for curriculum presentation in short-term, full-responsibility teaching are more demanding, and cover more aspects of teaching, than do the standards for individual lesson presentation. The first standard assumes the same form as one of the standards set for

the presentation of individual lessons. This is the pattern standard that requires evidence of favorable performance on the preponderance of teaching functions assessed in the sum of the lessons presented in the 2 to 5 days of full-responsibility teaching. The second standard for lesson presentation is also a pattern standard, and requires that over the course of the lessons presented in the period of full-responsibility teaching:

—variety in learning activities will be provided;
—variety in cognitive functions and levels in pupils will be exercised;
—variety in affective expressions will be employed in teaching; and
—positive feelings in pupils, such as excitement and interest, will be utilized in their learning.

Standards for Affect Management

The standards set for the management of affect in short-term, full-responsibility teaching take as their focus four dimensions of affective expression:

—teacher responses to instances of pupil affect;
—the management of pupil responses to instances of pupil affect;
—the anticipation of pupil upsets and disruptions, and their redirection; and
—the management of pupil upsets and disruptions when such occur.

Performance standards in relation to these dimensions of affective expression require that during the course of the full-responsibility teaching experience a student need only to perform effectively three of the four dimensions specified (any three will do), and that he or she needs to perform to this level on only one of the 2 or more days that he engages in full-time teaching. Such a standard reflects the view that the management of affect in a classroom is a complex matter, and that in an initial full-responsibility teaching situation, performance standards set for it should not be particularly demanding.

Attachment B

DEMONSTRATION CONTEXT: LESSON PREPARATION AND PRESENTATION

LESSON PLAN
(Attach to your Lesson Plan and Lesson Plan Evaluation Form)

Student's Name _____ *Lesson Number* _____

RECORD OF NEGOTIATION

APPROVAL TO NEGOTIATE THE LESSON WITH A SCHOOL SUPERVISOR

College Supervisor _____ *Date* _____

APPROVAL TO PREPARE A FORMAL PLAN FOR THE LESSON

School Supervisor _____ *Date* _____

* * *

CONTEXT DESCRIPTION

STUDENTS TO BE TAUGHT

School _____

Grade _____

Number _____

Special Characteristic: _____

CONTENT TO BE TAUGHT

Area _____

Expected Learning Outcomes _____

DATE(S) OF LESSON PRESENTATION _____

TIME(S) OF LESSON PRESENTATION _____

* * *

SUMMARY EVALUATION OF LESSON PLAN
(Obtain only after all elements of your plan have been evaluated)

I judge the plan as a whole to be of (circle one) →

	ACCEPTABLE QUALITY	OUTSTANDING QUALITY
School Supervisor		
College Supervisor		
Content Specialist		

* * *

Field Test Format 2
Experimental Elementary Teacher Education Program
Oregon College of Education
December 1972

LESSON PLAN EVALUATION FORM

Student's Name *Lesson Number*

Have the evaluators that check your plan initial each of the items listed that meets with their approval. If the treatment of an item is thought to be outstanding, have the evaluator draw a circle around his or her initials. Be sure to attach this sheet to your lesson plan.

ELEMENTS OF THE PLAN	*EVALUATORS OF THE PLAN*		
	School Supervisor	*College Supervisor*	*Content Specialist*
OBJECTIVES			
Are the learning outcomes expected from the lesson clearly stated?			
Are they appropriate and worthwhile outcomes, given the characteristics of the pupils to be taught?			
Are the indicators that are to be used as evidence of successful outcome achievement identified?			
Are the procedures to be used in obtaining evidence of outcome achievement identified?			
ADAPTING OBJECTIVES TO LEARNER CHARACTERISTICS			
Are there provisions for modifying the objectives of the lesson to meet individual pupil characteristics?			
SELECTING INSTRUCTIONAL MATERIALS AND PROCEDURES			
Are the instructional materials to be used in the lesson clearly identified?			
Are they appropriate to the learners to be taught and the learning outcomes to be achieved?			
Are the organizational and instructional procedures to be used in the lesson clearly identified?			
Are they appropriate to the learners to be taught and the learning outcomes to be achieved?			
EVALUATION			
Are there provisions for determining where pupils stand with respect to the desired learning outcomes of the lesson before it is presented?			
Are there provisions for feedback to pupils about their performance during the time the lesson is being presented?			
Are there provisions for determining where pupils stand with respect to the desired learning outcomes of the lesson after it has been presented?			
PLANNING NEXT STEPS			
Is there some indication in the plan of what would be done next with the pupils if the learning outcomes expected from the lesson materialize?			
Is there some indication in the plan of what would be done next with the pupils if the learning outcomes expected from the lesson did not materialize?			
MATCHING INSTRUCTOR, LESSON AND CONTEXT			
Does the lesson as planned appear to be feasible and appropriate to the school setting in which it is to be presented?			
Does the lesson as planned appear to be feasible and appropriate to the student who is to present it?			

DEMONSTRATION CONTEXT: LESSON PREPARATION AND PRESENTATION

SUMMARY EVALUATION OF LESSON PRESENTATION

Student's Name *Lesson Number* *Date*

We judge each of the lesson elements that follow to be of (circle one) →	*UNACCEPTABLE QUALITY*	*ACCEPTABLE QUALITY*	*OUTSTANDING QUALITY*
OBJECTIVES			
ADAPTING THE LESSON TO LEARNER CHARACTERISTICS			
QUALITY OF PRESENTATION			
Classroom Management (Transition, Termination, and Student Attention)			
Instruction (Materials, Procedures, and Organization)			
EVALUATION			
Preassessment			
Feedback during the lesson			
Feedback after the lesson			
Achievement of desired learning outcomes			
PLANNING NEXT STEPS			
THE FIT BETWEEN INSTRUCTOR, LESSON, AND CONTEXT			

School Supervisor *College Supervisor*

Student Presenter

Content Specialist *Student Recorder*

* * *

Field Test Format 2
Experimental Elementary Teacher Education Program
Oregon College of Education
December 1972

DEMONSTRATION CONTEXT: LESSON PREPARATION AND PRESENTATION

LESSON PRESENTATION EVALUATION FORM

Student's Name | *Content Area* | *Date* | *School or College Supervisor*

Lesson Element	*Evaluator's Judgment (check one)*			*The Behavior Of The Student and or Pupils That Led You To Make The Judgment You Have Made*	*Suggestions for Improvement*
	Yes	*No*	*Insufficient Evidence To Judge*		
OBJECTIVES					
Were the learning outcomes expected from the lesson clearly explained to the pupils?					
Were the learning outcomes expected from the lesson understood by the pupils?					
Did the pupils view the expected learning outcomes as appropriate and worthwhile?					
ADAPTING THE LESSON TO LEARNER CHARACTERISTICS					
Was the lesson carried out in a manner that reflected adaptation to pupil characteristics?					
QUALITY OF PRESENTATION					
Was the transition between the previous learning activity and the present lesson effective?					
Was the attention of the pupils maintained throughout the lesson?					
Were instructional materials, procedures, and organizational strategies used advantageously?					
Was the termination and "wrap up" of the lesson effectively handled?					

Field Test Format 2
Experimental Elementary Teacher Education Program
Oregon College of Education
December 1972

LESSON PRESENTATION EVALUATION FORM (continued)

Lesson Element	*Evaluator's Judgment (check one)*			*The Behavior Of The Student and/or Pupils That Led You To Make The Judgment You Have Made*	*Suggestions for Improvement*
	Yes	*No*	*Insufficient Evidence To Judge*		
EVALUATION					
Was the student's prelesson assessment of pupils essentially accurate?					
Was the feedback provided to pupils about their performance during the lesson adequate?					
Was the feedback provided to pupils about their performance after the lesson adequate?					
Were the desired learning outcomes from the lesson realized?					
Were there unexpected outcomes from the lesson that overshadowed the desired learning outcomes?					
PLANNING NEXT STEPS					
In talking with the student after the lesson was he or she able to identify next appropriate steps for pupils?					
THE FIT BETWEEN INSTRUCTOR, LESSON, AND CONTEXT					
Did the lesson as presented fit well into the ongoing instructional program of the school?					
Did the lesson as presented fit well the student who presented it?					

Field Test Format 2
Experimental Elementary Teacher Education Program
Oregon College of Education
December 1972

Attachment C

FIELD TEST FORMAT #3

COMPETENCY DEMONSTRATION CONTEXT: LESSON TEACHING

COMPETENCY CLUSTER I, PLANNING AND PREPARING FOR INSTRUCTION

_____
Student

_____ _____
Lesson Number Date

_____
Evaluator

The first formal demonstration of teaching competency in the elementary teacher education program at OCE takes place in the context of lesson teaching. Three lessons are taught for purposes of competency assessment, and three clusters of competencies are assessed in each lesson: Planning and Preparing for Instruction, Performing Instructional Functions, Performing Assessment Functions. A fourth cluster of teaching competencies, Displaying Pupil Achievement, is assessed in at least one of the three lessons. Performance standards and statements of procedure that accompany competency demonstration in the context of lessons are described in "The OCE Guide to Competency Assessment in Lesson Teaching."

THE FORMS ATTACHED ARE TO BE USED IN EVALUATING PLANS FOR EACH LESSON THAT IS TO BE TAUGHT

The Elementary Teacher Education Program
Oregon College of Education
Monmouth, Oregon
July 1973

COMPETENCY DEMONSTRATION CONTEXT: LESSON TEACHING

COMPETENCY CLUSTER I. PLANNING AND PREPARING FOR INSTRUCTION

(Attach to your Lesson Plan and Lesson Plan Evaluation Form)

Student's Name _____ Lesson Number _____

RECORD OF NEGOTIATION

APPROVAL TO NEGOTIATE A FORMAL LESSON PLAN WITH A SCHOOL SUPERVISOR

College Supervisor _____ Date _____

APPROVAL TO PREPARE A FORMAL LESSON PLAN

School Supervisor _____ Date _____

CONTEXT DESCRIPTION

STUDENTS TO BE TAUGHT

School _____

Grade _____

Number _____

Special Characteristics _____

CONTENT TO BE TAUGHT

Area _____

Expected Learning Outcomes

_____

_____

DATE(S) OF LESSON PRESENTATION

_____

_____

TIME(S) OF LESSON PRESENTATION

_____

_____

SUMMARY EVALUATION OF LESSON PLAN

(Obtain only after all elements of your plan have been evaluated)

I judge the plan as a whole to be of the following quality (enter the number in the box below your name that best describes your judgment as to the overall quality of the plan):

(1) MINIMAL

The elements of the plan are attended to at a level of detail, inventiveness, and care that suggests a *barely adequate* mastery of the knowledge and skill needed to present the lesson.

_____

School Supervisor

☐

(2) (3) ADEQUATE

The elements of the plan are attended to at a level of detail, inventiveness, and care that suggests a *sufficient mastery* of the knowledge and skill needed to present the lesson.

_____

College Supervisor

☐

(4) (5) OUTSTANDING

The elements of the plan are attended to at a level of detail, inventiveness, and care that suggests an *exceptional mastery* of the knowledge and skill needed to present the lesson.

_____

Content Specialist

☐

DEMONSTRATION CONTEXT: LESSON TEACHING

LESSON PLAN EVALUATION FORM

(Be sure to attach to your Lesson Plan and Summary Evaluation Form)

Student's Name _____ *Lesson Number* _____

DIRECTIONS TO EVALUATORS. When evaluating a lesson plan rate each of the BOLD FACE items that appear on this form that meet with your approval. If an item does not meet with your approval, the student is obligated to modify it until it does. All ratings are to be entered in the boxes provided. RATINGS FOR ALL ITEMS THAT APPEAR ON THIS FORM ARE TO BE BASED UPON PRODUCTS OF A STUDENT'S BEHAVIOR, RATHER THAN UPON BEHAVIOR PER SE. Each rating is to reflect one of the following judgments:

(1) **MINIMAL.** The element of the plan is attended to at a level of detail, inventiveness, and care that suggests a *barely adequate mastery* of the knowledge and skill needed to present the lesson.

(2)

(3) **ADEQUATE** The element of the plan is attended to at a level of detail, inventiveness, and care that suggests a *sufficient mastery* of the knowledge and skill needed to present the lesson.

(4)

(5) **OUTSTANDING** The element of the plan is attended to at a level of detail, inventiveness, and care that suggests an *exceptional mastery* of the knowledge and skill needed to present the lesson.

ELEMENTS OF THE PLAN	*EVALUATOR OF THE PLAN*			*COMMENTS*
	School Supervisor	*College Supervisor*	*Content Specialist*	
OBJECTIVES AND INDICATORS Are the learning outcomes expected from the lesson clearly stated? (Expected outcomes need not be stated in the form of "behavioral" objectives) Are they worthwhile outcomes, given the characteristics of the pupils to be taught? Are the indicators of outcome achievement identified?	☐	☐	☐	
INSTRUCTIONAL ACTIVITIES, MATERIALS, AND PROCEDURES Are the learning activities to be pursued in the lesson clearly identified? Are they logically related to the learning outcomes desired from the lesson? Are they appropriately sequenced? Are the instructional materials and procedures to be used in the lesson clearly identified? Are they appropriate to the learners to be taught and the outcomes to be achieved?	☐	☐	☐	

COMPETENCY DEMONSTRATION CONTEXT: LESSON TEACHING

COMPETENCY CLUSTER II; PERFORMING INSTRUCTIONAL FUNCTIONS
COMPETENCY INDICATORS: STUDENT BEHAVIOR

Student's Name | *Lesson Number* | *Date Presented* | *Time Presented* | *Supervisor or Content Specialist's Name*

DIRECTIONS TO EVALUATORS. When evaluating a lesson presentation rate each of the instructional functions (bold face items) that appear on this form. All ratings are to be entered in the boxes provided. RATINGS FOR ITEMS LISTED ON THE FIRST TWO PAGES OF THE FORM ARE TO BE BASED UPON STUDENT BEHAVIOR. These ratings are to reflect one of five possible judgments:

(1) **UNACCEPTABLE**
The behavior of the student in presenting the lesson reflects an *insufficient* level of thoroughness, inventiveness, and care in relation to the function.

(2)

(3) **ACCEPTABLE**
The behavior of the student in presenting the lesson reflects an *adequate* level of thoroughness, inventiveness, and care in relation to the function.

(4)

(5) **SUPERIOR**
The behavior of the student in presenting the lesson reflects an *exceptional* level of thoroughness, inventiveness, and care in relation to the function.

INSTRUCTIONAL FUNCTION	*SAMPLE INDICATORS*	*INDICATORS THAT SERVED AS A BASIS FOR JUDGMENT*
CONVEYING THE OBJECTIVES OF INSTRUCTION ☐ Was there an attempt to convey the objectives of the lesson to pupils?	Positive to Negative Examples • Objectives of the lesson are described • Steps are taken to insure objectives are understood • Reference is made to objectives during the lesson • Objectives are not stated • Objectives are stated once but not referred to again • Objectives are stated, but learning activities are not clearly related to them	COMMENT
ADAPTING INSTRUCTION TO CONTEXT ☐ Were the objectives of the lesson adapted to fit pupil responses to the lesson? Were the objectives of the lesson adapted to fit the characteristics of the instructional setting generally? Were instructional procedures and activities adapted to fit pupil responses to them? Were instructional procedures and activities adapted to fit the characteristics of the instructional setting generally? Were materials adapted to fit the characteristics of the instructional setting?	Positive to Negative Examples • Objectives are modified or differentiated for pupils on the basis of pupil response to the lesson • Objectives or procedures are modified to accommodate unexpected events; e.g., a snowfall or a fire nearby • Materials originally planned for use in the lesson are discarded or modified • Objectives, materials, or procedures are not modified, even though pupil response to them indicates they are inappropriate • Objectives, materials, or procedures are not modified to accommodate unanticipated events	COMMENT

PERFORMING INSTRUCTIONAL FUNCTION—PAGE 4

THE RATING THAT IS TO BE PROVIDED ON THIS PAGE IS STILL TO REFLECT ONE OF THE FOLLOWING JUDGMENTS:

(1) **UNACCEPTABLE**

No more than half the pupils being taught behaved in ways which indicated the instructional function was being carried out effectively.

(2)

(3) **ACCEPTABLE**

Three-fourths or so of the pupils being taught behaved in ways which indicated the instructional function was being carried out effectively.

(4)

(5) **SUPERIOR**

All or nearly all of the pupils being taught behaved in ways which indicated the instructional function was being carried out effectively.

INSTRUCTIONAL FUNCTION	*SAMPLE INDICATORS*	*INDICATORS THAT SERVED AS A BASIS FOR JUDGMENT*
MANAGING UNEXPECTED EVENTS Were potentially disruptive events effectively managed? ☐	Positive to Negative Examples • Unexpected visitors or unusual events do not cause undue disruption in learning activities • When a child is disruptive the response of the student to him does not add to his disruptiveness or cause it to spread to others • A child who disrupts tends not to be disruptive a second or third time • When a child is feeling angry or upset or afraid the student's response to him does not intensify his feelings, or cause them to spread to others • Negative indicators are reflected in behavior that is the reverse of the above	COMMENT

COMPETENCY DEMONSTRATION CONTEXT: LESSON TEACHING

COMPETENCY CLUSTER III, PERFORMING ASSESSMENT FUNCTIONS
COMPETENCY INDICATORS: STUDENT BEHAVIOR

Student's Name | *Lesson Number* | *Date Presented* | *Time Presented* | *Supervisor or Content Specialist's Name*

DIRECTIONS TO EVALUATORS. When evaluating a lesson presentation rate each of the assessment functions (bold face items) that appear on this form. All ratings are to be entered in the boxes provided. RATINGS FOR ALL ITEMS THAT APPEAR ON THIS FORM ARE TO BE BASED UPON STUDENT BEHAVIOR. These ratings are to reflect one of five possible judgments:

(1) **UNACCEPTABLE**
The behavior of the student in presenting the lesson reflects an *insufficient* level of thoroughness, inventiveness, and care in relation to the function.

(2)

(3) **ACCEPTABLE**
The behavior of the student in presenting the lesson reflects an *adequate* level of thoroughness, inventiveness, and care in relation to the function.

(4)

(5) **SUPERIOR**
The behavior of the student in presenting the lesson reflects an *exceptional* level of thoroughness, inventiveness, and care in relation to the function.

ASSESSMENT FUNCTION	*SAMPLE INDICATORS*	*INDICATORS THAT SERVED AS A BASIS FOR JUDGMENT*
ASSESSING LEARNING BEFORE THE LESSON ☐ Were steps taken before the lesson began, or early in the lesson, to identify where pupils stood in relation to the learning outcomes expected from the lesson? If such an assessment was made, was it carried out with reasonable efficiency and accuracy? Were objectives, activities, or procedures in the lesson adapted to reflect the information that came from the assessment?	Positive to Negative Examples • A pretest over the learning outcomes expected from the lesson is administered prior to the lesson • The children's regular teacher is asked for her judgment as to where the pupils stand relative to the expected learning outcomes • One of the first activities undertaken in the lesson is an "informal" assessment of where the children stand relative to such outcomes • Whatever the form of assessment, it is carried out in a relatively short period of time and without undue attention being called to it • Information coming from the assessment is acted upon in designing the lesson • Neither a formal or informal assessment is made of pupil learning before the lesson is undertaken • If an assessment is made it is inefficiently or ineffectively carried out; for example, it requires too much time, gains too much attention, or the information that is obtained is inaccurate or in some other way unsatisfactory • Information is obtained but there is no indication that it is acted upon	COMMENT

COMPETENCY DEMONSTRATION CONTEXT: LESSON TEACHING

COMPETENCY CLUSTER IV, DISPLAYING PUPIL ACHIEVEMENT
COMPETENCY INDICATORS: PRODUCTS OF A STUDENT'S BEHAVIOR

Student's Name | *Lesson Number* | *Date Presented* | *Time Presented* | *Supervisor of Content Specialist's Name*

DIRECTIONS TO EVALUATORS. When evaluating a lesson presentation rate each of the achievement display functions (bold face items) that appear on this form. All ratings are to be entered in the boxes provided. RATINGS FOR ALL ITEMS THAT APPEAR ON THIS FORM ARE TO BE BASED UPON PRODUCTS OF A STUDENT'S BEHAVIOR, RATHER THAN UPON BEHAVIOR PER SE. These ratings are to reflect one of five possible judgments:

(1) **UNACCEPTABLE**

The products of the student's effort to display learning outcomes related to the lesson reflect an *insufficient* level of thoroughness, inventiveness, and care in relation to this teaching function.

(2)

(3) **ACCEPTABLE**

The products of the student's effort to display learning outcomes related to the lesson reflect an *adequate* level of thoroughness, inventiveness, and care in relation to this teaching function.

(4)

(5) **SUPERIOR**

The products of the student's effort to display learning outcomes related to the lesson reflect an *exceptional* level of thoroughness, inventiveness, and care in relation to this teaching function.

ACHIEVEMENT DISPLAY FUNCTION	*SAMPLE INDICATORS*	*INDICATORS THAT SERVED AS A BASIS FOR JUDGMENT*
DISPLAYING PRELESSON AND POSTLESSON ACHIEVEMENT ☐ Were the results of the assessment before the lesson recorded and displayed in such a way that they were easily interpreted? (Please attach the display if one has been prepared; sample display forms are found in the "OCE Guide to Competency Assessment in Lesson Teaching") Were the results of the assessment after the lesson recorded and displayed in such a way that they were easily interpreted? (Please attach the display if one has been prepared; sample display forms are found in the "OCE Guide to Competency Assessment in Lesson Teaching")	Positive to Negative Examples • Results of at least the postlesson assessments of learning outcomes are recorded and displayed so that anyone looking at the information can easily determine where each pupil stands with respect to the learning outcomes expected from the lesson • Postlesson learning scores are summarized to show the proportion of pupils who achieved the learning outcomes expected from the lesson, and the proportion that did not • Prelesson and postlesson assessment information is not recorded or displayed • Postlesson achievement data are displayed but are not summarized to reflect the proportion of pupils who achieve the learning outcomes expected from the lesson, or the proportion of students who fail to do so	COMMENT

Estimating Costs of a Competency-Orientation

By

BRUCE R. JOYCE

How do we build a competency-oriented certification and training system? Personally, I could never go through all the necessary complicated steps, although there are people here who could. I'm much more interested in trying to find out what teaching skills will pay off; whether they have to be idiosyncratic; whether anybody can even train us all in the same ones. It's conceivable that we all may not be capable of the same ones; maybe in the long run we will have to do our own thing as teacher competence, in the end, may turn out to be personal—or it may be external and trainable. Also, teaching is elusive. You can measure 2,000 reasonable-sounding competencies and not have caught the essence of teaching. So far the research hasn't validated many competencies, but I'm not prepared to give it up.

To determine competence, we also have to face some rough questions. For example, we're not even sure that as teachers we should be playing the roles we do. I keep critizing the role of the self-contained classroom teacher mostly because it nearly killed me. I had to go into college teaching to keep alive, but I simply couldn't teach six things to 36 kids who were that different from each other. I didn't have the capacity for the job, and I had to become a professor in order to survive. I'm not sure that I want to spend any time finding competencies for the present roles teachers play. We might do much better to design better schools and find out what competence we need to make them work.

I know nothing about costs, OK? I couldn't cost-account a major program, but last summer, some of the New York State Education Department people asked me to help them estimate how much a competency program would cost to develop. I agreed to do it, and I wrote them the following paper.

Largely because of the massive research and development effort entailed if competency-based training and certification are to be achieved, it has been suggested that development and implementation be centralized. An alternative to accommodate local needs and bring about a broad base of participation in development is that the competency-orientation should be required and relatively small local consortia (probably made up in each case of a few school districts and colleges) should determine competency criteria, develop programs, and implement the programs and the certification procedures. The latter option is attractive because it promises to involve so many people and to insure relevance to local school needs.

The centrally mandated system is inacceptable on both political and substantive grounds: politically because it would give considerable power to a few persons (competency-oriented certification and training is much more likely to affect who will actually be permitted to teach than course-based certification and training, however centralized), and substantively because not enough is yet known about the identification of teaching competencies to permit anyone to develop a widely mandated set of competencies for any particular category of teachers with any great confidence that those competencies will stand up.

Centering development around many local centers is acceptable as a process, but is not without serious problems. The more local centers there are, the more difficult it will be to organize teams of sufficient expertise to develop really strong training and certification systems. Major universities, working with strong school districts, have had trouble doing this. Yet, strong development teams are necessary. High quality in a competency-orientation is essential because both certification and training procedures will be more powerful and mistakes (such as emphasizing trivial competencies) will be magnified.

Many districts and smaller colleges are unaccustomed to freeing personnel for development, but a sustained effort by a large team is essential if the task of identifying competencies, organizing training and assessment, and implementing certification procedures is to be done effectively.

Without coordination and quality control, broad participation in the program development could be an endless and ineffective process.

Purpose

In this paper an attempt is made to identify the tasks of creating a competency-based teacher training and certification system and to estimate the costs of completing the tasks adequately. In addition, a base is provided for estimating the effect on costs of several possible options for organizing the process of development and implementation. This can be used as a base for estimating cost when several possible options for organizing the process include broad local input into the system. To estimate basic technical costs, an estimate is made of the cost of development by a compact, independent team of experts who would simply accomplish the technical tasks of producing a prototype training-certification system. This option (referred to as process option A) is not a realistic choice, but yields a fairly reliable cost estimate from the technical point of view, which permits an estimate of the cost if several possible organizational processes were employed; e.g., strictly local development (process B), local development with central technical assistance and development services (process C), and a general state system with provision for local options (process D).

Source of Cost Estimates

How do we arrive at the costs of the tasks and subsequently, the costs of the process options? There are three sources of our estimates. One is the experience of the Bureau of Research Teacher Education Project, especially the costing procedures included in their feasibility studies. The second is the cost of developed performance-based materials such as those in "Materials for Modules" (appendix B). These two sources can help us determine the base costs represented in process option A (the least desirable process option in terms of creating a statewide training-certification system). The third source is the least reliable and represents estimates of the probable increase in the base cost due to the greater complexity and duplication of effort of the other process options.

Technical Tasks and Process Options

Certification and training are interrelated parts of the same system which serves as the basis for preservice teacher education, inservice teacher education, and the certifying and diagnostic procedures related to both of them.

The purpose of the development of competency-oriented systems is to unite preservice education, initial certification, inservice education, and continuing or extended certification around a system of teaching competencies which form the goals of preservice education, the standards for initial and extended licenses, and the basis for diagnosis of performance of inservice teachers and prescription of inservice training. It is probably not possible or desirable to separate licensing or training at the preservice or inservice levels under the competency orientation.

The entire system for competency-oriented certification and training depends on the creation of four interrelated storage and retrieval systems plus the organizational and communications networks necessary to create and implement them. The creation of these systems and networks constitutes the technical side of the development process. The creation and management of the organization necessary to create the systems constitute the process side. Both technical and process sides present options which affect costs greatly.

The Technical Side

The technical side of competency-based certification and training involves the creation and validation of four systems. These are as follows:

A storage and retrieval system of the specifications of teacher competency.

Since teaching is complex the number of competencies which are likely to be specified is large indeed. The Bureau of Research Models averaged between 2,000 and 3,000 competency specifications and these represented efforts by single unified institutions or small consortia rather than statewide consortia of diverse institutions and representation. It seems reasonable to suppose that the number of competencies will increase as the political base for establishing them is broadened. Hence, when teacher associations, representatives of school administrations, the public and students all contribute, as well as expert teams from universities and state departments of education, the number of potential competencies of a teacher will be large and the process of identifying the most important competencies may be complex.

A storage and retrieval system of mediated instructional systems and agent-mediated components designed to produce the competencies.

This storage and retrieval system represents the means of teacher education. The number of items stored as instructional systems will be equal to or somewhat greater than the number of competencies which are specified. Also, the extensiveness of each instructional system is much greater than the specification of any competency. Thus, the production of software to fill this system will be an extensive task. Put more simply, it will be difficult to specify critical competencies, but developing the program elements to achieve them will be even more complex.

A system of assessment devices designed to determine the effects of the instructional systems, and agency-mediated components and to measure the competencies specified in system A.

Without an assessment system competency-based certification would be impossible as would be the assessment of the effectiveness of program elements. Properly organized, the assessment system provides a diagnostic profile of the teacher candidate and provides also the means for tracking his progress to determine when certification should be granted. In the case of inservice teachers, it provides the means for diagnosing the state of their competency and relating them to the instructional systems to be used in inservice education.

A management system for interrelating subsystems 1, 2, and 3 above.

The magnitude of the three other systems requires the use of a contemporary management system for diagnosis, prescription, tracking and progress, and providing feedback to teacher candidates, teachers, and program and licensing managers. Without a complex automated management system individualization or personalization would be impossible in a program and so would be implementation, for the complexity of specifications and training devices is so great as to defeat any presently available option for program control. Without a modern information system chaos would surely ensue at the point of implementation. It is possible to imagine a statewide management system which individual teacher training institutes would use to guide them in the identification of competencies, and they might use a central system to withdraw instructional systems and assessment devices for their particular training program.

The research and development effort needed to produce the four interrelated systems is enormous. They represent the cost of the substance of competency-based teacher education and licensing.

The Process Side

If it is desired to have participation by all relevant groups in the creation of the competency-based teacher education and certification system then suitable communications networks and organizations have to be set up to permit teachers, students, teacher organizations, representatives of the public, subject matter experts, experts on teaching, and representatives of state departments and colleges to participate in competency identification, the selection of instructional systems, and the creation of the assessment systems. The process of creation can reasonably be divided into three levels: specification of the systems, development, and implementation. It is possible to imagine a state or nationwide network providing participation by all relevant groups in the specification level of the creation of the basic system. Development will probably have to be organized at a few major centers, funded to bring together talent to create the operating systems. It is possible to imagine a network stretching across the state so that many local units might contribute something to the development. However, development is an exceedingly complex task, especially if the products are to be tested as they are developed. Thus, a few major centers probably will do most of the actual production of the instructional systems and testing devices. However, broad-based groups can evaluate the products, and a representative group can monitor the entire operation.

Minimum, General, Statewide Competencies and Extended Local Competencies

It is possible to use some of the critical competency specifications for minimum statewide certification—these could represent minimum standards for preservice and inservice training. A much larger number of specifications might be prepared representing local needs. These would be used by individual teacher training centers for training purposes and for determining local certification in addition to state certification. Imagine, for example, a network of field centers representing the local school districts, higher education institutions, community representatives, teacher associations, etc. Suppose that there are about 20 of these in various parts of the state. They produce competency specifications and agree on those which will be used for minimum specifications for any given type of teacher (elementary, secondary, special education, etc.). Each of the local centers, however, would produce specifications which were deemed important by

the people in that center but which were not selected by statewide representatives as essential for statewide certification. These additional specifications nonetheless could be stored in the specifications storage system and instructional systems could be designed to achieve them and assessment systems to assess them. Thus a complete statewide system, if it is to take care of local needs, would have to be much larger than the system which would produce only state requirements. Clearly the greater the participation of interested parties and the more numerous the local centers, the more costly the creation of the system. From the point of view of organization and communication and also from the point of view of the size of resulting system participation and with provision for local options costs will increase.

This cost would be nothing like the cost of asking each local unit or center including several local units to create their own competency-based teacher education and certification system. The cost of building any strong, competency-based program with local options will be $10 to $12 million (and that is probably the minimum which should be anticipated). It would cost $5 to $7 million for each local unit to create its own system. Thus if there were 10 systems in the state each developed by local units, the cost would be in the neighborhood of $60 to $70 million. It would seem far more reasonable in cost to create a statewide system that had provisions for special local needs and interests.

Technical Tasks and Process Options

The four tasks involved in the creation of a competency-based program, plus the development of a certification system, represent the five tasks required to activate a statewide competency-oriented training and certification system. The cost of each of these will vary considerably depending on the process options which are selected. In table 1, four process options are depicted over the five tasks.

Process A (developing one complete prototype system) is the least costly option in terms of dollars, but has the disadvantage that its statewide acceptance would not be high, in all likelihood.

Process B (development of complete system by several local centers) would be very costly. The cost of each complete system would be higher than that of a system developed by an "expert" team (I estimate that the cost would be 50 percent higher) and that would be multiplied by the number of local centers.

Process C (local development supported by a few development centers) would be somewhat less costly than B because the development centers could reduce duplication of effort.

Process D (development of a statewide system with local options) would be much higher than A but much lower than B or C and would result in an acceptable statewide process and plan. It appears to be the best cost option.

Process A is least costly because most compact. It is useful, however, for cost estimates because it can yield a base cost which can be multiplied by factors representing the increased complexity and duplication of effort required in the other options.

Process B results in several comparable systems. It gives local needs the fullest play, but the use of nonexpert teams escalates the cost of each system so that each local system would be very expensive and the cost

Table 1. Cost Factors of Technical Tasks by Process Options

System	*A. Prototype by Expert Team*	*B. Local Consortia as Focus*	*C. Local Consortia Plus Development Assistance*	*D. Statewide System With Local Options*
1. Competency System	1	1.5 (no. of consortia)	1.25 (no. of consortia)	2
2. Instructional System	1	1.5	1.25	2
3. Assessment System	1	1.5	1.25	2
4. Management System	1	1.5	1.25	2
5. Certification System	1	1.5	1.25	2

within the state would be the cost of each system times the number of local systems.

Process C would be a bit less than B due to some centralization of effort.

Process D, providing for wide local participation but central coordination to reduce duplication of effort, might only require about twice the investment of process A.

The Base Cost of the Technical Tasks

The Bureau of Research Models, especially the feasibility studies, provided cost estimates of many of the basic technical tasks. These are supplemented by the actual costs of developing performance-based training materials. The Florida State University and University of Wisconsin estimates provide data especially relevant to the types of activities which New York appears to require.

General Costs

The Florida State team attempted to identify the various costs of starting and implementing a competency-based teacher education program that would provide comprehensive clinical training. That is, they were dealing with the professional components rather than the liberal arts components and the others that might contribute to the education of a teacher. They planned about 300 program units. They estimated about $2¼ million to develop the units, about $75,000 to develop and test the entry diagnostic system, about $400,000 to develop and equip the computer management system, and about $200,000 to carry on a faculty training program, making a grand total of a little less than $3 million.

My personal view is that this estimate is conservative. For example, the faculty training program is not costed very effectively by years. It is worthwhile noting, however, that they probably assume that many of the faculty would be persons who are involved in the development and who would, in the course of developing the materials, train themselves to carry on the kinds of activities they would need to engage in as faculty members. It is worth noting, that they did not expect to implement a program in 1 year or 2 but rather expected to take 5 years for the development and implementation process.

I think this is a very fair estimate of time. Developing, testing, and integrating really significant, competency-based instructional systems will be a time-consuming and very expensive task. Furthermore,

Bruce Joyce

if one were to go to a competency-based licensing, one would not want to do so until he had determined the general effects of a competency-based teacher education program. In addition to the time for "start up" that was estimated by Florida, it might take 4 or 5 years before enough people had been through the program observed as teachers to lay a data base for revising the certification procedure. Thus, to move from the present teacher education program structure to a competency-base could be somewhere between 5 and 10 years, probably in the upper end of that range. Process options would not affect time much, because development of materials is the chief consumer of time and the process options would affect only the other tasks.

The above costs did not include competency identification, which averaged a bit over $100,000 for each of the Bureau of Research studies. Since the Bureau of Research teams generally feel that the specification would have to be reworked before development could begin, probably another $100,000 would have to be invested in this task.

The Cost of Data Processing and Management System

It is not possible to imagine competency-based education on a large scale unless it is supported by a computerized management system. The eventual programs may be somewhat less complicated than those envisioned in the Bureau of Research models, but units in

a competency-based program will amount to a very large number. (Probably more than 2,000, still under the average number in the Bureau of Research programs.) In relating 500 to 1,000 teacher candidates to a program containing that many elements requires information access, the coordination of support materials, and the coordination of faculty of very great complexity.

The cost of creating such a management system will be considerable. The University of Wisconsin did a very thorough job of identifying the elements of a management system and the costs of maintaining a staff sufficient to operate the system. They thought out a system which would have the capacity of handling over 700 students in an on-campus setting. They estimated an annual cost of around $712,000 to maintain the system (about $1,000 per student.) The rental of hardware for the system would be in the neighborhood of $350,000, and the cost of the data processing staff about the same. However, such a management support system would have the capability of serving a great many more students at relatively low additional cost. If a state the size of New York decides to move into competency-based education, it would seem wise to develop one or two prototype management systems. If carefully designed they could accommodate the program units for a number of programs of different types. This effort would reduce substantially the cost of developing and maintaining management systems and institutions throughout the state. However, it is very hard to imagine that even the greatest amount of sharing among the centers for teacher education would reduce the cost of maintaining a management system to much under $500 per student. The system, however, would permit an individualization of instruction, program planning, and assessment far beyond the capacity of any present teacher education program. Also the system envisioned by Wisconsin is a multimedia system using computer system instruction, instruction through motion picture, television tapes, audio tapes, and programed units providing a variety of instructional modes which very few instructional settings presently offer. Thus the utility of the management system is general —it handles a large number of program elements for a large number of students and increases the type as well as number of instructional options.

On the Requirement of an Automated System

It is worth noting that none of the Bureau of Research models was conceived without the assumption that it would be possible to operate a competency-based program without an automated management system. To relate 500 students to 2,000 instructional units in such a way that students have instructional options, are assessed and made aware of results, and relate to program options in terms of developed competence, a management system is simply necessary.

The expectation should be that part of the development of the competency-based system involves the creation of management systems. No program plan should be accepted that does not include the provision for the development as such an automated system.

Base Cost Estimates

Under process option A, the costs of the tasks would seem to break down as follows:

Developing the Competency System		$200,000
Developing Instructional Materials (including Assessment Devices)	minimum	$2,500,000
Developing Diagnostic System		$75,000
Developing Management System		$400,000
Annual Cost of Maintaining Management System (to serve one program)		$750,000

These can be multiplied by the factors in table 1 to develop very rough estimates of costs for the other process options.

The experience of developers in recent years adds some specificity to the cost estimates and suggests cautions.

Development of Specifications

The cost to the United States Office of Education of having nine models developed by relatively expert teams in a very short period of time (about 8 months — probably not a sufficient period of time to do the job right) was over $3½ million. Relatively inexpert teams working over a longer period of time will probably be much more costly. The cost could be reduced by having a statewide organization in which state and local teams take responsibility for certain areas and the whole effort is coordinated. Local participation, especially to lay down initial preferences and to monitor the results, could be encouraged.

The Development of the Training System

Thus, the training system area is where the real cost actually begins. The Bureau of Research project estimated between $5 and $7 million to develop their system, using a single team within a consortium or within

a single institution of higher education linked to local school districts. The broader the base of participation and the more complex the model of teacher education, the higher will be the cost.

The cost of the mini-courses developed at the Far West Laboratory is illustrative. The Far West Laboratory uses an extremely efficient development team which creates and tests its product using standard research and development procedures. The teams are well trained and stay together for long periods of time. They do not have responsibility for a wide diversity of products but concentrate on the particular type of instructional system known as the mini-course. The cost of development of each single mini-course providing about 30 hours of instruction to a teacher on one teaching skill is over $100,000. At Teachers College, Columbia University, a series of six instructional systems designed to teach three teaching skills and three strategies was developed at a cost of a little over $100,000, or about $16,000 per instructional system.

It is probable that a comprehensive teacher training program at the preservice level will contain at least 200 instructional systems (as in Florida's estimates). If that is the case and the development cost of $15 or $16 thousand found by the Columbia team turns out to be an accurate one, the estimates from the Bureau of Research feasibility studies are about on target.

This will be the cost of developing one model of teacher education by one centrally organized team. As indicated earlier, the greater that development is decentralized (process options B, C, D) the greater will be the cost unless the teams specialize and do not attempt to build the entire range of possible instructional systems.

What might be done would be to develop a master plan of development and subcontract pieces of the development (process D). However, it has to be remembered that there are very few places presently in the state where much developmental productivity could be expected for some time to come and it would take centers a while to tool up. Therefore, it seems wise to make the assumption that a few places in the state would have to be funded for major development efforts.

Further Notes on the Assessment System

The picture is somewhat different with respect to the assessment subsystem. The developers of the instructional systems could create many of the assessment measures for specific competencies as a part of their developmental effort. This would probably be the most efficient way of developing the specific measures and was the procedure proposed by the Bureau of Research team. However, in addition to the measures of specific competency, as they would relate to the instructional systems, one needs measures of the general competency of the teacher or situations in which he can bring together a group of competencies and perform in a more effective way. It was recommended in several of the Bureau of Research proposals that teaching laboratories be used for this; that situations be set up in which a teacher could be brought together with a small group of students. He would teach for an hour or two, while his ability to set objectives and teach, and the gains in pupil knowledge were assessed at the same time. Some kind of assessment system like this would probably be desirable. It is unlikely that rating blanks or rating systems of observing and rating teacher class behavior in the teacher's own classroom will prove to be feasible. For more than 50 years persons have been trying to develop such measures, with notable lack of success.

Saving Money by Sharing Materials

The technical nature of the competency-oriented system makes it unusually dependent on the creation of competency-based software. Specifying the competency necessary for any particular teacher role, building appropriate instructional systems, creating reliable assessment systems, and implementing a program all require extensive research and development efforts and major changes in procedures. Statements of competency, instructional systems, and assessment devices are all either wholly (as in competency specifications) or partly expressed on paper, film, television tape, etc. Without a very large development effort, there will be no implementation of CBTE.

Under some circumstances, a solution to the need for development might be to fund one or two agencies to create complete systems and disseminate them.

In this case, however, the actual implementation must be closely allied to local needs and local personnel must participate substantially in the whole process of determining directions for change, defining teacher roles, and selecting and implementing training systems.

To support local efforts without either controlling them or engaging in a ruinously costly duplication of development costs in every local agency, a large national storehouse of competency-based products should be developed. From this storehouse, local agencies can

draw much of what they will need to build their competency-based training system, while developing the remainder locally. Fortunately, agencies across the Nation have been producing materials which can get the storehouse off to a good start.

The area of reading instruction provides a good example. There presently exist a number of approaches to reading and a number of systems for acquainting teachers with them and training competencies relevant to those approaches. A local agency can draw from these in order to build the reading component of a program to train preservice or inservice teachers.

This situation is duplicated in several other areas.

The developers represent a range of agencies, many of which are wholly or partially supported by Federal funds:

1. Research and development centers and regional laboratories are available.
2. Curriculum projects in many areas, including ubject areas, approaches to the education of young children, special education, correction, career opportunities, etc., are in existence.
3. N.C.I.E.S. projects, including the Teacher Centers, Texas Education Agency, Protocols projects, Florida State, are funded.
4. Teacher Corps programs have included development officers for the last 2 years and their technical assistance projects have contributed products and processes.
5. Colleges and consortia which are in the process of reaching for the competency orientation have produced many materials.

The result is a beginning to a national storehouse—provided that the products are brought together, made visible, and demonstrated in preservice programs and inservice teacher centers. More products are needed, to be sure, and many need transformation and testing, but a considerable start has been made and should not be lost. Many agencies are presently operating from scratch unaware of the considerable array from which they could draw.

What is necessary to make the storehouse useful? First, potential users need to know what exists in the storehouse and the potential use of each product. (Presently the Florida Catalog of Competencies, the Florida Catalog of Competency-Based Materials, and "Materials for Modules," prepared at Teachers College for the Teacher Corps, list many materials.)

Second, users need to know how materials actually work—what an instructional system or assessment device can be expected to achieve and what it takes to use it.

Third, some materials need further development to make them adaptable to many local situations—their transportability needs to be improved.

Fourth, users need to experience competency-based programs, and Teacher Centers can operate with products largely drawn from the storehouse. This is extremely important. The storehouse has now reached the point where energetic designers can, by using the products of others and developing some products themselves, actually operate programs and Teacher Centers which would be at least three-fourths competency-based.

The surprising fact is that there is presently available much more software, especially in terms of instructional systems, than could possibly be used in a 2-year, full-time, teacher preparation program. Not all of the material is of high quality—and development has been missing in some important areas—but careful selection from the storehouse, combined with local development, will enable program implementation fairly soon, provided that the available materials are disseminated and demonstrated and if program models and Teacher Centers using the storehouse are also established as demonstration centers.

The Nature of the Storehouse

Most of the products are in the form either of instructional systems or smaller units, generally called modules. Generally they consist of a competency specification, a set of activities and supporting mediated material, and an assessment device. Materials are available in at least the following areas. (See following page.)

Imagine, for example, a program to prepare teachers of the social studies.

It might include competency in:	And adapt the following materials:
(1) the study of teaching	Parsons' Guided Self-Analysis or Mini-course on Flanders System
(2) basic educational psychology	(Grinder program)
(3) basic instructional design	Vincet (Popham)
(4) curriculum alternatives in social studies	I.I.U. (Far West Laboratory) Teachers College Units
(5) basic teaching skills	Mini-courses Teachers College Units
(6) basic teaching strategies	Presentational skills (General Learning Corporation) Teachers College Units

ARRAY OF PRODUCTS

Areas of Competence	Types of Competence			
Study of Teaching	Introduction to Systems for Studying Teaching	Basic Skills for Studying Teaching	Skills in Analyzing Teaching of Self and Peers	Application of Skills in Feedback to Teachers
Basic Teaching Skills	A vast variety of teaching skills are embodied in instructional systems. (These are often expressed in jargon, as Structuring, Modulating Focusing, Reinforcing, Set Induction, Explaining, etc.)			
Models of Teaching	A considerable variety of teaching strategies are available. (These can be in specific content areas, in Wallen's instructional systems in the reading area, Ivey's counseling, and mini-courses in arithmetic teaching; or they can relate to curriculum strategies, as the instructional systems to teach teachers the strategies of IPI, or how to implement the Engelmann-Becker approach to early childhood education.)			
Instructional Decision-Making Skills	Setting Objectives	Selecting Approaches	Selecting Evaluation Devices	
	There are many types of systems for instructional design, and competency-oriented devices to teach them have proliferated.			
Diagnostic Skills The Study of the Student	Creativity	Self-Esteem	Achievement	Sociability
Circular Decision Making	Reading	Social Studies	Science	Mathematics
	(Most of the development in this area consists of modularized curriculum courses developed by local programs. However, products like the Information Units from the Far West Laboratory are professionally packaged. The local products vary in quality but include some excellent materials.)			
Instructional Design	Architecture	Staff Utilization	Curricular Design	Technology
	(These areas are relatively untouched by developers.)			
Educational Philosophy	Personalists	Social Idealists	Academics	Cybernetics
	(There are some modular methods courses in these areas, but much needs to be done.)			
Educational Psychology	In this area there are a variety of packaged courses. Developers tend to differ in point of view, so content varies.)			

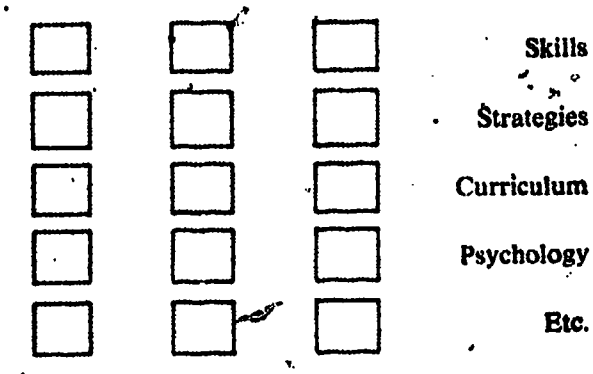
Using the already-developed materials (including competency specifications, instructional materials, and assessment devices), one could easily organize a faculty to implement a competency-based program to prepare teachers of the social studies. Most teacher education agencies would probably prefer to create some of their own materials, but to develop all they would need would probably require an investment of $400 or $500 thousand, whereas if they draw from the national storehouse the cost becomes feasible.

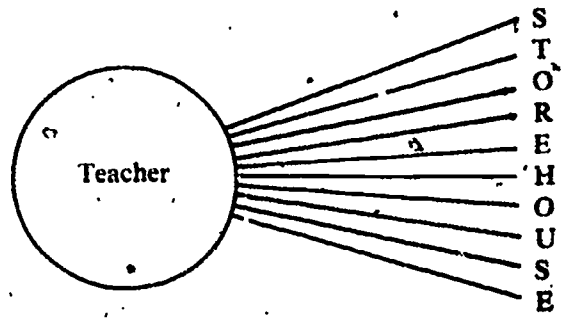
A Teacher Center

To build a Teacher Center would involve much the same process. A Teacher Center should offer teachers three types of service:

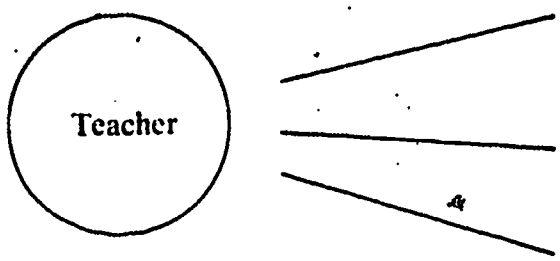
1. *Teachers Option.* When the teacher selects competencies according to his interest or self-diagnosis.
2. *External Diagnosis Option.* When the teacher's peers or supervisors determine that he has a need by analyzing his teaching.
3. *Curriculum Thrust Option.* When a team, school, or agency inaugurates a new program requiring new competencies.

To create such a center one needs to offer teachers a great many options which can be selection on any of the three service bases described above. For example, instructional systems could be classified on a basis somewhat like the Array of Products and made available with a microteaching laboratory. Thus:





Teacher Centers could be designed around the roles of the teacher. A good many of the existing products are suited to the various roles. For example:



General Competencies	*Role Related Instruction*
Study of Teaching	Skill Builder (IPI Training System) (SRA System in Reading)
Instructional Design	Productive Thinker (Synectics Material)
Basic Teaching Skills	Community Leader (Teachers College Units)
	Counselor (Micro-counseling Skills)
	Academician (McCrel Units, University of Texas Units, Teachers College Units)

More than enough material presently exists in the storehouse to permit the creation of Teacher Centers such as this. Coupled with the types of workshop centers used in the English Teacher Centers, they provide a great many possibilities for teacher self-training.

Cost and Quality

Even if a local program built all its own staff, I think CBTE is dirt cheap even at the level of cost I have estimated. I think it is extremely inexpensive. What is really expensive is what we are doing now. We pour hundreds of millions of dollars into ineffective teacher education programs every year. That's almost pure waste. The inefficiency and waste of the present system is just terrible.

A great deal of the cost of development will be in-kind costs. View a teacher education program as a developing nation. You have to develop capital somehow and there are a couple forms of capital: one is ideas; the other, software. You invest faculty time in order to get these. We simply need to put a greater proportion of staff costs into development than we have.

What bothers me about the stance that it will be easy and cheap is when development simply becomes modularizing the present stuff of teacher education. That *is* easy *and* cheap but it doesn't result in much improvement.

I have never personally experienced a major university (maybe Houston, maybe Toledo) that was capable of doing the whole development job by itself. The National Consortium is essential for just pooling the talent to do the job. For example, without the Mini-courses produced at the Far West Lab and other materials produced by dozens of others, I wouldn't dream of the type of program we run at Columbia although we build a lot of instructional systems ourselves. We must share materials if we are to keep costs in bounds and draw on each other for program ideas. CBTE should be a national development effort with a myriad of local variations.

Bibliography

I. Models for Elementary Teacher Education

The Florida State University, A Model for the Preparation of Elementary School Teachers (OE-58018; two volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

Michigan State University, Behavioral Science Elementary Teacher Education Program (OE-58024, three volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

Northwest Regional Educational Laboratory, A Competency Based, Field Centered, Systems Approach to Elementary Teacher Education (OE-58020, three volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

Syracuse University, Specifications for a Comprehensive Undergraduate and Inservice Teacher Education Program for Elementary Teachers (OE-58016, two volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

Teachers College, Columbia University, The Teacher Innovator: A Program to Prepare Teachers (OE-58021), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

University of Georgia, Educational Model Specifications for the Preparation of Elementary Teachers (OE-58019), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

University of Massachusetts, Model Elementary Teacher Education Program (OE-58022), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

University of Pittsburgh, A Model of Teacher Training for the Individualization of Instruction (OE-58018), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

University of Toledo, Educational Specifications for a Comprehensive Elementary Teacher Education Program (OE-58023, two volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1968.

University of Wisconsin. Elementary Teacher Education Project (OE-58025, four volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1969.

II. Feasibility Studies

The Florida State University, A Feasibility Study of the Florida State University Model for the Preparation of Elementary School Teachers, Washington, D.C.: U.S. Office of Education Bureau of Research, 1969.

Michigan State University, Feasibility Study: Behavioral Science Teacher Education Program, Washington, D.C.: U.S. Office of Education Bureau of Research, 1970.

Ohio Consortium, The Feasibility of Educational Specifications for the Ohio Comprehensive Elementary Teacher Education Program (Phase II), Washington, D.C.: U.S. Office of Education Bureau of Research, 1970.

Oregon College of Education, A Plan for Managing the Development, Implementation and Operation of a Model Elementary Teacher Education Program, (three volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1970.

Syracuse University Protocooperative, A Study of the Feasibility of the Refined Syracuse University Specifications for a Comprehensive Undergraduate and Inservice Teacher Education Program for Elementary Teachers, Washington, D.C.: U.S. Office of Education Bureau of Research, 1969.

University of Georgia, The Feasibility of the Georgia Educational Model for Teacher Preparation—Elementary, (three volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1970.

University of Massachusetts, A Feasibility Study on the Model Elementary Teacher Education Program (Phase II), Washington, D.C.: U.S. Office of Education Bureau of Research, 1970.

The University of Toledo, The Feasibility of Educational Specifications for the Ohio Comprehensive Elementary Teacher Education Program (Phase II), Washington, D.C.: U.S. Office of Education Bureau of Research, 1969.

University of Wisconsin, Wisconsin Elementary Teacher Education Project Feasibility Study, (two volumes), Washington, D.C.: U.S. Office of Education Bureau of Research, 1970.

III. Other Sources

Cruikshank, Donald R., et al., The Ohio State University Analysis of the Nine Comprehensive Elementary Teacher Education Models (CETEM), Washington, D.C.: U.S. Office of Education, National Center for Educational Research and Development, 1970.

Engbretson, William E., Analysis and Evaluation of Plans for Comprehensive Teacher Education Models, Washington, D.C.: U.S. Office of Education, 1968.

Joyce, Bruce R., "Variations on a Systems Theme: Comprehensive Reform in Teacher Education," Interchange, Vol. 1, No. 3 (1970), pp. 83-95.

Klatt, Judith and Walt Le Baron, System Development Corporation, A Short Summary of Ten Model Teacher Education Programs, Washington, D.C.: U.S. Office of Education, National Center for Educational Research and Development, 1969.

Le Baron, Walt, System Development Corporation, Systems Analysis and Learning Systems in the Development of Elementary Teacher Education Models, Washington, D.C.: U.S. Office of Education, National Center for Educational Research and Development, 1969.

Proceedings of 22d Annual Georgia Teacher Education Conference: TEACHER EDUCATION MODEL, Journal of Research and Development in Education, Vol. 2, No. 3 (Spring 1969).

A Reader's Guide to the Comprehensive Models for Preparing Elementary Teachers, Washington, D.C.: ERIC Clearinghouse on Teacher Education and the American Association of Colleges for Teacher Education, 1969.

System Development Corporation, Analytic Summaries of Specifications for Model Teacher Education Programs, Washington, D.C.: U.S. Office of Education, National Center for Educational Research and Development, 1969.

Funding Competency-Based Teacher Education

By

HERBERT HITE

Competency-based teacher education costs differ from the costs of standard teacher education because the two programs differ. There are differences in the required performances of students and also in the roles of faculty. Program materials are different and so are assessment procedures. The sources of funds for CBTE, however, are not different from the sources for traditional, or standard, programs. The same sources of revenue for higher education which are always in short supply must implement the CBTE programs. Therefore, the analysis of costs for CBTE is most meaningful when these costs are expressed in terms of funding formulas for other higher education programs.

Among the program variables of CBTE which are reflected in differing costs from standard programs are the following: [1]

1. Instructional materials, units, and techniques are individualized. If the criterion of CBTE is the success of the individual student of teaching, then CBTE materials must allow for different rates and different options by these individual learners. These materials constitute a higher cost factor than standard program materials.
2. For the most part, the CBTE faculty work in a one-to-one relationship with students—as counselors, evaluators, explainers, analysts. The changed role of faculty has a higher cost than faculty in the traditional lecturer role.
3. The process of admission to CBTE appears to be a more complicated and hence a more costly process than admission to standard programs. Prerequisites for success, which should constitute admission criteria, seem to be more difficult to assess than the usual examination of students' grade-point averages.
4. Complex management systems may be a significant cost item.
5. Students in the CBTE programs are usually full-time interns or observers or work over varying periods of time with individualized study materials. Full-time students rather than number of classes are the bases for measuring costs. This is not necessarily a higher cost factor, just different.
6. In CBTE, school personnel have a significantly larger role in the instructional process and the cost of their participation is relatively higher.
7. More instructional time is involved in assessment of student progress, and this process is usually more complex than grading in courses. Evaluation costs are usually higher.
8. Decision making regarding all phases of the CBTE program involves not only college personnel but also school and community members. The large amounts of these people's time which is needed are a significant cost factor, and no funds have been available for this function in the past.
9. Specialist or consultant services to assist state agencies in program approval may result in additional costs for CBTE.
10. Competency-based programs leading to advanced certification will be centered in school districts, and the development and administration of these new programs will be an additional cost factor eventually.

In a study of costs for the new CBTE certification standards in the state of Washington, which was requested by the state's legislature, two CBTE programs were analyzed.[2] The study suggests that for the time being, the following factors are additional costs of CBTE relative to standard teacher education programs in that state:

1. Add 50 percent to the costs of the standard teacher education to account for the variable of individualized mode of instruction.
2. An additional 50 percent will be required for developing each new CBTE program.
3. An additional 50 percent will be needed to pay for released time of school personnel involved in consortium arrangements—policy making, program planning, securing program evaluation, etc.

[1] These 10 factors are described in a paper given for the Teacher Corps in June 1972, "The Economics of Competency-Based Teacher Education," Herbert Hite, Research Foundation of the State of New York, Albany, N. Y., 1972.

[2] One of the programs analyzed in the cost study was described in the paper previously cited.

The recommendations to the Washington State Legislature are for the colleges and universities to receive double funding for new CBTE programs and the school districts to receive funds for consortium activities.

Some of the costs for CBTE are the same if the program includes one student or 1,000 students. These costs include the development of policy councils, the certification process, the development of basic competencies and assessment processes, and a basic set of individualized materials. Some costs vary according to the number of students: the number of faculty per student, the number of laboratories or public school classrooms, the number of sets of materials. Some costs represent quality judgments, such as the number of options to be made available to students, the ratio of faculty to students, the degree of refinement of structional modules and assessment procedures.

Obviously, a good deal of the difference in costs of CBTE and standard programs depends upon decisions as to the nature of specific CBTE programs. For example, the definition of competency can make a difference in program costs. If competency is defined as a set of teaching performances, then the focus will be upon highly refined and tested instructional devices to elicit teaching behaviors. If the emphasis is upon products of teaching; i.e., pupil achievements, then the emphasis will be upon the field setting for working with learners and evaluating the outcomes of this activity.

The CBTE model at Western Washington State College is based upon product criteria. The candidate is required to demonstrate competency by bringing about "appropriate" changes in the behavior of elementary or secondary pupils. The major components of the model are an entry phase, in which the candidate meets admission criteria, a knowledge phase in which he acquires a repertoire of information about the substance of what he will teach, a laboratory in which he acquires and demonstrates knowledge and skill about teaching and learning principles and applies these skills, and a practicum in which he demonstrates his ability to apply principles of teaching and learning in a variety of situations. The laboratory and practicum constitute two quarters of full-time study and practice in a teaching center.

The program depends upon close working relationships in the laboratory and practicum among the students, the professors, and cooperating teachers who go through a special training program to prepare as field supervisors. The basic essentials of the program are evaluations of entry into the program, diagnosis of knowledge necessary to succeed in the laboratory, a test of instructional competency as an entry into the practicum and modified T.I.C.'s as evidence of mastery of the basic model and hence certification. Study materials in the laboratory consist of 52 modules and assorted readings, films, filmstrips, tapes, observation assignments, and mini-teaching arrangements. The students design their own study programs leading to the T.I.C., which is the basic measure of competency, by negotiating learning contracts with the clinical professors. Study contracts vary.

After 4 years of partial success and occasional failures, the clinical faculty has arrived at a working model. The costs of this working model are heavier than the costs for standard teacher education at Western. The cost studies which have been completed suggest that even after initial years of program development have been completed, there are these factors to add to the costs of the standard programs at Western:

1. About 15 percent additional is needed to fund extra support costs, such as visits by campus-based faculty to the teaching centers, and special individualized learning modules.
2. About 25 percent additional is needed for the continuous development of the program components—evaluation instruments, adaptations of modules, processes for orienting experienced teachers to the model, etc.
3. About 25 percent is needed for the administration of each teaching center, including negotiations with school personnel concerning consortium arrangements.

This is a modest proposal for extra funding compared to other suggestions for funding CBTE. The Western model is still developing, but it is based upon modest resources. Compared to the Elementary Teacher Education models, funded in 1968 by USOE, this is a Model T compared to a brand new Lincoln Continental. The present program does produce the required demonstrations of competency, however, and there are considerable affective gains over the standard program in the views of faculty and students.

Still, even a modest program needs somewhere between a 50 and a 100 percent increase in funding for teacher education. Where does the funding come from? In the real world, deans of education do not confront the legislature or the boards of regents with demands for so many extra millions to fund a new program. Colleges and universities have only a few alternatives.

These are the possible alternatives a college has for finding the additional resources for CBTE:

1. The college might maintain the existing level of faculty and instructional support funding but limit the enrollment of students into the program.
2. The college might require more time or credit hours by students who were candidates for a certificate under the CBTE plan, thus increasing either tuitions or student credit hours and consequent funding.
3. The faculty might seek an adjustment in the formula for support of programs. Field-based or laboratory programs might be funded at twice the level of campus-based programs.
4. The college faculty might seek outside funding —grants or fellowships.
5. Schools might contribute by sharing the costs through support of a complementary inservice CBTE program.
6. Colleges might develop graduate programs, with their higher rates of funding, as adjuncts to the competency-based undergraduate program.

In Washington, master's level instruction in state institutions is funded at 2.3 times the level of upper-division student instruction. Clearly this level of funding would provide what the faculty at Western have defined as their need for CBTE programs. This adjustment would require action by the state legislature or a major and radical change in programing teacher education within the college. Neither possibility seems likely.

Outside funding is possible, but not very dependable as a base for a continuing extra program cost. The college could reduce enrollment and maintain the existing level of faculty and support programs, but this is politically impossible at Western, although this may be possible at other institutions. A few school districts will share costs. In fact, one school district reportedly has a budget of $150,000 for advanced certificate programs under the CBTE model. These school funds, however, are usually restricted to payments to school personnel for released time. This leaves the college administrator where he always is—with the same old system to try to manipulate. Basically the game is to generate sufficient credit hours to secure the necessary entitlement for faculty and instructional support.

The CBTE program at Western operates in four teaching centers. Each center consists of a team of clinical professors, students, and cooperating teachers. In a center with four clinical professors, here is a strategy for generating the necessary student credit hours and getting the necessary work accomplished.

Two clinical professors will have eight to 10 new students each quarter. As the students remain for two quarters, this means that the instructional load for each of the two professors will be 18 full-time students, at the undergraduate level. These two professors between them generate three and one-third full-time faculty equivalent entitlements.

Herbert Hite

One clinical professor takes on half as many students, or a total of nine, which is about two-thirds of a full-time equivalent faculty load. He is also responsible for the administrative matters of the center, including negotiations with school personnel.

One clinical professor will direct 25 part-time graduate students in a four-credit practicum. The students are teachers and volunteer to act as cooperating adjunct faculty in the CBTE program. The credits they earn may apply to an advanced certificate. Their practicum is concerned with development of the processes for working with students in the program, including evaluation criteria and techniques. The major responsibility of this clinical professor is program development. The 25 students in the four-credit graduate practicum generate one full-time equivalent faculty load.

The four clinical professors, then, generate five F.T.E. units, which entitles the center to the needed instructional materials and services which are additional to those required in campus-based programs. Also, two of the professors earn student credit hours in the process of administering and developing the CBTE program.

The four clinical professors may modify their own assignments. They are jointly responsible for 45 full-

time undergraduates and 25 part-time graduates, however they divide the responsibilities of the teaching center. These figures are only illustrative.

A variation on this model is a center in which the undergraduate intern, in his final quarter of demonstrating competency, relieves an experienced teacher, who becomes a full-time master's candidate. The intern is backed up by a first-quarter laboratory student who acts as a teaching assistant part time. Both are closely supervised by the clinical professor. In this model, instead of supervising 18 full-time undergraduates, the clinical professor is responsible for four full-time graduate students, four interns, and four laboratory students, in four classrooms.

These figures and load descriptions represent a few of the ways in which the CBTE faculty may generate the amount of credit hours needed to support the costs of CBTE which are above the costs of standard programs. The experience at Western indicates that these kinds of loads are feasible, given the nature of the CBTE program which has been developed at that institution.

Summary

Competency-based teacher education is more expensive than the standard programs. The variables within CBTE programs which affect costs are many, and, additionally, there are many kinds of value judgments about programs which will affect the level of funding needed for a given CBTE program. Funding for a modest approach to CBTE, which is largely field-based, is possible within the usual formulas for generating support for higher education. The experience of clinical professors indicates that the credit hours generated by full-time undergraduate students and part-time graduate students (cooperating teachers) are sufficient to fund CBTE. State officials could simplify the funding problem, however, by supporting CBTE as a "high-cost" college program, equivalent to graduate study or to study in the other professional fields such as health sciences or law.

It seems clear from initial experiences with CBTE that costs are greater than with regular programs. How much greater depends upon a number of judgments which are made by the CBTE program designers.

The high cost factor which is common to all CBTE programs is the individualized mode of instruction, which is necessary if teacher candidates are to demonstrate competency criteria. Other expenditures, particularly expenditures for program development and for management, can vary wildly from CBTE program to CBTE program.

The cost decision may hinge upon the way the faculty define competency. If competency is defined in terms of specific teaching behaviors which are assumed to be related to changes in pupils, then the modules and other training materials will need to be refined to the point that it is possible to predict a given level of performance on the part of the consumer of those modules. It follows from that decision that teaching modules and strategies must be highly refined. Costs can be very high to achieve this end.

Competency may be defined, however, as changes in the performance of the teaching candidate's clients. In that case the specific modules or strategies will be effective inasmuch as they assist an individual candidate to implement growth on the part of the candidate's pupils. The emphasis in program development will then be upon the field setting, the education of cooperating school personnel, the evaluation of the products of teaching. The costs for these components of CBTE are more likely to be absorbed within the ongoing school program for pupils and the inservice education for teachers.

If the CBTE faculty choose to design all their own learning modules, the costs will be relatively high—they could be astronomical. The cost factor seems comparable to the costs of developing programed instruction or instructional systems. There are choices to be made concerning the quality levels of the CBTE program; e.g., the ratio of faculty to students, the number of alternative instructional strategies, etc.

New state certification standards which specify competency or performance by candidates also require that the new teacher education programs be managed by a consortium of agencies, including not only colleges or universities but also school administrators, professional associations, and the public. Released time for teachers to participate fully in training programs and in policy decisions is a new cost for teacher education.

The realistic level for funding CBTE might be to compare teacher education to graduate education or to other high-cost programs such as in the laboratory sciences.[4] The rationale for shifting the funding from a level comparable to academic college instruction to that of funding graduate or laboratory science instruc-

[4] The higher level of funding per student does not mean that the total cost of teacher education necessarily would be higher: preparation agencies *could* limit enrollment in their CBTE programs.

tion is that CBTE is like other high-cost collegiate instruction. CBTE is a highly individualized program and should be supported in the same way as other collegiate programs which are conceded to require a highly individualized approach.

Ultimately, the preparation of teachers under a competency-based model will need sustained financing in the same way as teacher education programs are presently funded. Specific CBTE programs must earn college credits. The credits must meet the quality standards other college credits must meet. The credits must be sufficient to provide for the necessary time of faculty and for support programs. The additional funding needed to implement the individualized CBTE programs must come through the existing system of higher education. Therefore, the fate of CBTE seems to rest with the decision makers who allocate funds to colleges and universities, and specifically CBTE will depend upon a more generous formula than is now used to fund undergraduate teacher education.

Assessment of Teacher Performance: What is Involved? What is the Cost?*

By

BEATRICE A. WARD

When the term "performance-based" is used as a descriptor for teacher training and/or teacher certification, certain essential conditions apply. What the teacher does is as important as what he knows. During training, skill development receives equal or greater attention than knowledge acquisition. At the end of training, mastery is used as the evaluative criteria rather than a list of courses completed and grades received. Information about the teacher's performance is obtained throughout the training period and subsequently in the operable classroom in order to determine whether the specified teaching skills are being acquired and used. Student performance also is studied in order to determine whether teacher use of certain skills relates to the level and quality of student learning.

Within such a framework, assessment of teacher performance is multi-faceted. It is conducted for many different reasons, focuses upon a wide variety of teacher and student behaviors, and involves a diversity of individuals and institutions. Inquiry into several of these facets of assessment is the purpose of this paper. The discussion centers around choices that may be made regarding assessment and the costs associated with particular options. It builds upon the information obtained by the Teacher Education Division at the Far West Laboratory for Educational Research and Development during 5 years of developing and testing performance-based teacher training materials.

Why Assess Teacher Performance

Assessment of teacher performance can be undertaken for a variety of purposes. Among them are the following:

- To determine the average type of performance that can be expected of experienced or inexperienced teachers at a local, regional, state, or national level
- To identify the types of teaching skills a majority of inservice or preservice teachers perform well and those which they need to improve or acquire
- To establish a minimum acceptable level of performance for success as a teacher; to identify those teaching skills that relate to successful student performance
- To determine what an individual's level of performance is at the time he enters inservice or preservice teacher training
- To establish that an individual has mastered one or more teaching skills as a result of training

Two types of data are associated with these forms of assessment. Assessment conducted to determine average teacher performance, identify how well teachers perform certain specific skills, or establish a minimum acceptable level of performance builds upon information regarding teaching in general. Samples of teaching and teachers may be used as a data base. The information obtained serves as a guide for policy decisions such as the selection of the content (skills and knowledge) to be included in a preservice and/or inservice training program and the specification of the skill and knowledge requirements for certification. General information of this type also aids in determining how extensive an inservice retraining program should be; i.e., the potential number of teachers needing training in a given skill and the geographic areas in which particular types of training are needed.

On the other hand, the data base for planning an individual teacher's training program and/or verifying that a particular individual has acquired certain specified teaching skills requires detailed information regarding that person's performance. Sampling procedures cannot be used. Rather, a profile of skill usage along with a judgment as to the quality of use must be available for each individual teacher. The specificity of the profile may vary depending upon whether the pattern of skill use will be used to outline a training

* Invited paper, Multi-State Consortium on Performance-Based Teacher Education, New Orleans, Louisiana, February 27, 1973.

sequence or verify competence prior to certification. Pupil learning outcomes may or may not be part of the profile depending upon the inservice or preservice status of the teacher. But, regardless of the information included, the assessment process must help each individual teacher identify his areas of strength and weakness and plan a training program that will improve his teaching performance.

Assessment of the individual teacher or assessment of a sample of the teacher population, therefore, evolves from the intended use of the assessment information. Policy decisions can build from a different data base than training decisions. Obviously, given sufficient time and financial resources, assessment of individual teachers could serve policymaking as well as training purposes. However, since time and money are limited, study of a representative sample of teachers appears to be the most cost/effective means for obtaining the data needed to make policy decisions. A variety of techniques may be used to arrive at the teaching sample. For example, an adaptation of matrix sampling (Husek and Sirotnik, 1968; Lord and Novick, 1968) might provide a means for obtaining information about how a broad sample of teachers' use of a wide range of skills at the same time keeping costs similar to those associated with an indepth study of a small sample.

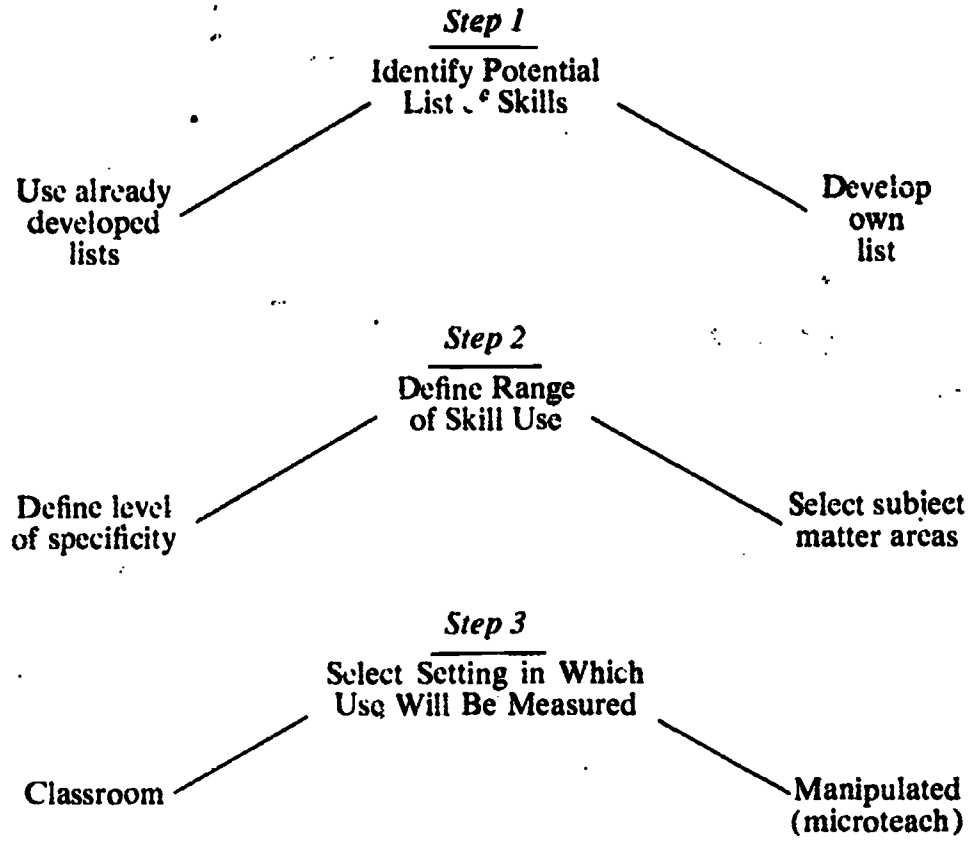
Based upon these guidelines, the ultimate cost of teacher assessment is determined, in part, by the purposes for which the findings will be used. A first step toward designing and financing asn assessment effort is to outline the questions the data are to answer.

What Teaching Performance Will Be Assessed?

Determining what teaching performance will be assessed also involved several steps (see figure 1). The first is to identify the potential list of skills to be included in the assessment effort.

Figure 1

Steps in Selecting Teaching Performance To Be Assessed



Identifying skills. The diversity of knowledge and skills encompassed within the teaching act has been documented by numerous experts in the field: i.e., Smith, 1971; Houston, 1972; Joyce & Weil, 1972. Yet, to date, research has shown few positive relationships between teaching skill and student outcomes (Rosenshine, 1971). As a result, selection of the skills to be assessed, for the most part, consists of combining what little evidence exists regarding teacher effects with theoretical and practical opinion regarding desired teacher performance.

Several efforts to identify critical teaching skills have already been undertaken by individuals and agencies such as the developers of the models for elementary teacher education (Elementary Teacher Education Models, 1969), the designers of competency-based teacher training programs (Dodl, 1972), and the developers of performance-based teacher training materials (the Far West Laboratory). In each instance, a thorough review of the literature related to a particular aspect of teaching is undertaken, inclass observations are made to verify that certain skills are used, and experienced teachers, school administrators, parents, and others are asked to judge the relative importance of the skills.

As suggested by Joyce (1973), persons responsible for teacher assessment may opt to use these already developed skill lists or they may choose to develop their own. If the latter option is taken, the costs presented in table 1 indicate the expenditures in personnel and other operating expenses such as travel and purchase of supplies and materials that may be required. The figures are based upon costs incurred by the Teacher Education Division of the Far West Laboratory while identifying critical teaching skills in the areas of pupils' language development, mathematics tutoring, independent learning, and higher cognitive questioning. On the average, skill identification cost $14,670 for each set of skills.

The efficacy of using and/or building upon existing skill lists is apparent given these costs. Since teaching incorporates multiple sets of skills of the type identified by the Laboratory, the total cost of skill identification could be expected to exceed $100,000 if started from the beginning. This would be the cost even if the search were to be limited to available research information, and theoretical and experiential opinion. If further research to identify relationships between teacher and pupil performance were included, the costs would be much higher. Thus, if assessment costs are to be kept within reason, the skills to be assessed should be taken from the skill lists being compiled as part of the broad performance-based teacher education effort. Each assessment agency should not attempt to develop its own unique skills list.

Range of Skill Use. The second step in determining what teacher performance will be assessed is to define the range of skill use to be studied. By this, I mean the level of specificity at which the skill will be assessed and the variety of subject matter areas in which the skill samples will be taken.

Many areas of teaching can be described in terms of general performance levels as well as specific categories within these levels. For example, the general teaching skill of asking higher cognitive questions can be divided into the more specific skills of asking analysis, synthesis, and evaluation questions (see table 2). Within any assessment process, whether it is for policy-making or training purposes, measures of this skill probably would focus upon the general performance level. Previous research indicates that higher cognitive questions generally represent less than half the questions asked by teachers (Gallagher, 1965; Davis & Tinsley, 1967; Guszak, 1967). In order to set training priorites

Table 1. Costs of Identifying a Set of Teaching Skills

	Skill Area				
	Pupils' Language Development	*Math Tutoring*	*Independent Learning*	*Higher Cognitive Questioning*	*Mean*
Personnel Costs	$ 5,584	$10,068	$ 4,953	$ 5,981	$ 6,647
Direct Costs	6,938	11,836	6,156	7,163	8,023
TOTAL	$12,522	$21,904	$11,109	$13,144	$14,670

Table 2

Higher Cognitive Questioning*

General Skill Level: Use higher cognitive questions

Specific Skill Level:

1. Use three types of analysis questions.
   —questions requiring students to think of motives or causes
   —questions requiring students to make inferences
   —questions requiring students to find evidence to support generalizations, interpretations, or conclusions
2. Use two types of synthesis questions.
   —questions asking students to make predictions
   —questions asking students to develop solutions to problems
3. Use three types of evaluation questions.
   —questions asking students to take a stand on a controversial issue
   —questions asking students to judge truth or validity
   —questions asking students to judge beauty or worth

and/or verify competence in this area, therefore, the single measure of percentage of higher cognitive questions asked is adequate. If a teacher (preservice or inservice) asks more than 60-70 percent higher cognitive questions, he is performing much better than the average teacher. If he is using considerably less than 50 percent higher cognitive questions, he probably would benefit from training. Given this general piece of assessment information, many decisions can be made. It is only after a decision to design and instigate training has been made that the specific skill levels become important. Learning to ask analysis, synthesis, and evaluation questions can help increase a teacher's overall use of higher cognitive questions. Specificity at the training stage is essential. For assessment purposes, general measures of skill use can provide adequate information and at the same time keep measurement efforts and costs within reasonable limits.

As outlined in figure 1, defining the range of skill use to be assessed also includes selection of the subject matter areas in which teacher performance will be studied. Again, this is a decision that affects assessment costs. Single examples of teaching are less costly to obtain than multiple examples.

For some teaching skills, the decision is obvious. Skills related to teaching reading should be assessed in the context of a reading lesson. Likewise skills specific to the teaching of mathematics need to be measured while mathematics is being taught.

Other skills are more general in nature. Probing, for example, can be used in any subject area. It employs the same type of teacher-pupil interaction regardless of the content of the lesson. A teacher's use of probing in any subject area probably would be representative of his use of the skill in general.

Still other skills differ in use depending upon the context in which they are applied; for example, response to pupils' errors in reading probably differs from response to errors in science.

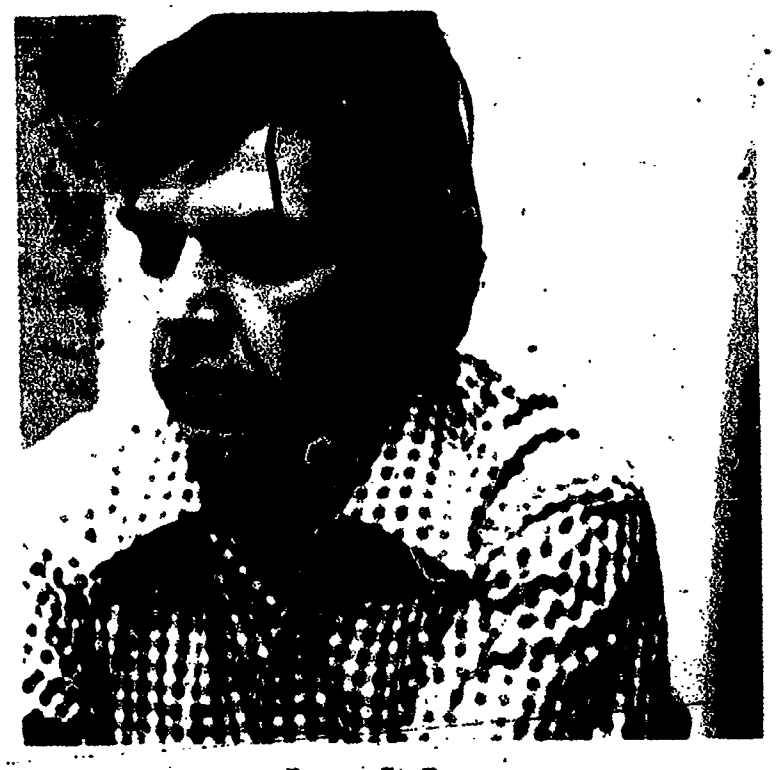
The problem within an assessment effort is to determine which subject areas are most likely to offer the best opportunity for teachers to exhibit their use of each particular skill. During the initial phases of assessment, inquiry probably will be limited to generalizable skills or to specific content area skills. Because of data collection complexity, information regarding skills that apply to several content areas but differ in their use within each area will follow later.

Performance Setting. The third step in selecting the teaching performance to be assessed is to select the setting in which skill use will be measured. Based upon Turner's (1972) six criteria of teacher performance, inclass as contrasted with manipulated examples of behavior may be considered; the most common manipulated example being a microteach lesson. The salient question to be considered is whether a reasonable sample of skill can be obtained in a microteach setting.

Since much teaching occurs in small group or one-to-one tutorial situations, and critics of the classroom (such as Silberman, 1970) contend that large-group

* Taken from Mini-course 9, Higher Cognitive Questioning, developed by Far West Laboratory.

Beatrice Ward



James Steffenson

instruction should seldom take place, a small-group setting, e.g., a microteach lesson, may be an appropriate information base for studying a high proportion of teaching skills. The disadvantage of such a performance sample is that few student outcomes of any consequence could be expected to occur in a single 10-20 minute lesson. Thus, if study of teacher effect upon student learning is to be included in the assessment process, at least some examples of inclass performance should be included. It should be noted, however, that given the present state of the art of assessment, researchers (Smith, 1971; Herbert, 1971) question the appropriateness of using student outcomes to measure teacher performance. Further, since the restricted environment of a microteach lesson, whether live or on videotape, also facilitates concurrent scoring of a number of teaching skills, this instructional setting warrants serious consideration as an assessment tool.

Cost of Assessment

The Teacher Education Division of the Far West Laboratory has conducted a number of studies of teacher change resulting from minicourse training. In these studies, an assessment procedure has been used that includes

- —assessment of already identified skills (skills were identified during development of training materials),
- —assessment of both general and specific skill use (e.g., both percent of higher cognitive questions and use of analysis, synthesis, and evaluation questions were assessed),
- —assessment within a certain prescribed content area,
- —assessment within a microteach setting.

Using this procedure, microteach lessons were recorded on videotape and critiqued at some later date by trained observers. Only limited, if any, measures of student performance were obtained.

Sample costs for conducting the assessment are presented in table 3. They are based on the study of five teachers because this is the number of teachers who can conveniently be scheduled into a microteach facility during a 9:00 a.m. to 3:30 p.m. school day. The average cost per teacher for five teachers is $180. Since critiquer training, research design, and data preparation costs would not be repeated with a larger sample of teachers, the cost to assess 30 teachers would be $2,160; or $72 per teacher.

If the skills to be measured and the lesson content were carefully selected, such an assessment effort could provide a large amount of information about teacher performance. It could be used for many of the purposes outlined earlier in this paper.

Who Will Do the Assessing?

Assessment of teacher performance generally will be conducted by three types of individuals.

1. *Teachers.* Peer observations (one teacher observing another and vice versa) can be used to study a wide variety of teaching skills.

Table 3.
Cost of Assessing Teacher Performance Using a Microteach Setting*

Collecting Performance Sample		
5 1/2-hour videotapes @ $15/each	$ 75.00	
7.5 hours of research intern to monitor collection of video tape examples (@ $9,450/year + benefits)	50.00	
		$120.00
Critiquing Performance		
Critiquer training;		
1 hour training per behavior		
12 behaviors × 2 critiquers × 1 hour		
24 hours × $3.50/hour	84.00	
12 hours of research intern to do training	77.00	
Critiquers to score recorded lessons		
12 behaviors = 6 passes to score		
6 passes @ 30 min. each = 3 hours per tape		
3 hours × 5 tapes = 15 hours @ $3.50 per hour		
15 hours @ $3.50 per hour × 2 critiquers	105.00	
		266.00
Research design and data interpretation		
1 week research design planning @ $17,000/year (Note: If more teachers were involved, this cost would be dispersed over the entire population)	434.00	434.00
Data preparation		
Keypunch — ½ day clerk/analyst at $9,000/year + benefits	40.00	
Run and analyze — 1 day	40.00	
		80.00
TOTAL COST (Assessment of 5 teachers)		$900.00

* Based on assessment of 5 teachers.

2. *Critiquers*. Analysis of audio and/or video tape recordings of teaching (e.g., a microteach lesson) may be done by specially trained critiquers. They also may be teachers but more often they will be graduate or undergraduate students in education hired to carry out a particular critiquing task.
3. *Inclass observers*. Whenever an assessment program demands high accuracy in recording teacher inclass performance, specially trained individuals will be needed. Inclass scoring of teacher performance requires simultaneous monitoring of multiple variables and instantaneous recognition of the skill(s) to be assessed. Considerable training and practice must be completed in order to achieve this degree of observer skill.

Information obtained through peer observations will be the least accurate form of assessment data. Nonetheless, when the assessment is being done to identify inservice teachers' training needs, this type of information is sufficient to establish a tentative skill profile.

In our work at the Far West Laboratory, we tested a form of peer assessment in the responsive skills area. Six observations were made (in this case, four were done by teacher peers and two were completed by the

teacher's students). Gross information was obtained regarding such skills as teacher response to pupil ideas, teacher use of positive and/or negative responses to student performance, and teacher use of nonverbal reinforcement. Given the observation findings, the teachers were alerted to skill areas in which they needed to improve. Decisions could be made regarding where to begin the inservice training program for that group of teachers. The gross information received from the peer observations allowed the training to proceed with some degree of individual focus.

Critiquer analysis of audio or video tape recordings of teaching samples provides more precise information about the skills used. The data are more reliable than peer observations. Many skills can be scored from a single teaching sample by rerunning the recording several times. On the other hand, since the teaching situation being studied is manipulated, questions must be asked regarding the generalizability of these findings to the teacher's regular performance in the classroom. The best time to use this form of assessment may be at the end of training when a teacher needs to be sure that he has acquired and is using a particular skill or set of skills and when nuances in skill use need to be pursued.

Overall, the most comprehensive observation procedure is to have highly skilled observers score teacher and pupil performance in the classroom setting. As noted above, this requires the observer to identify specific skills at the moment they occur. Teacher-pupil interaction in an on-going classroom setting cannot be rerun to check the observer's scoring. Further, since the observer can code the use of only a small number of skills at a time, several observations may be necessary in order to code the same number of skills that could be scored in a single 30-minute video taped lesson. Inclass observations appear to be most useful when pupil outcomes are to be assessed as well as teacher performance, when a particular teaching skill is expected to occur only in a natural setting, when data on teacher performance must be analyzed as soon as they are obtained, and when equipment to record teaching samples is in limited supply. Extensive inclass observations are not essential to five of the six assessment purposes listed earlier in this paper.

The cost of assessment depends upon both the type of performance to be measured and the assessment procedure to be used. Peer observations are the least expensive; inclass observations the most costly unless large amounts of equipment must be purchased to install critiquing of recorded lessons. Table 3 presented an example of the costs associated with assessment of microteach samples of teaching. Estimates of the cost of trained observer scoring of the same number of teaching skills in the classroom would be higher than the $72 per teacher projected for the microteach sample. Because of this increased cost, inclass observations probably should be used only when no other assessment method will provide the needed information.

Developing the Teacher's Skill

A basic assumption of teacher assessment is that once data are available regarding a teacher's performance, training will be offered to improve that teacher's competence. Using his skill profile to identify areas of strength and weakness, the teacher may waive training in one skill area to emphasize improving another. To be successful, a program that allows these options incorporates training packages that cover diversity of skills. Within a performance orientation, each of these packages generally will include models of the skills to be acquired, opportunities to practice the skills, and evaluative feedback regarding the teacher's performance during practice.

As noted by Joyce (1973), development of such training materials is costly. Time and effort, represented by dollars, must be spent to identify the skills to be included in the package, develop the initial form of the materials, test the materials to be sure teachers acquire the specified skills, and revise the materials based upon testing results. Table 4 presents cost data for the design of one such type of training package, the minicourses developed at the Far West Laboratory. The data represent average costs for each development task based upon the design and testing of four minicourses.

The merit of building a training system that uses as many already developed materials as possible is obvious given such development costs.

Table 4. Cost of Developing Training Materials

Development Task	*Cost*
Conceptualization and skill-identification	$ 14,670
Product development	28,962
Product testing (one test with 30–50 teachers)	46,279
Revision after testing	12,627
TOTAL	$102,538

An additional cost factor that should be considered as part of a complete assessment cycle is the cost of providing training. Table 5 lists the estimated costs for a 5-week training sequence that includes microteach-

Table 5. Cost of Training*

Start-up costs		
Materials purchase		
16mm filmed models of teaching skills		$ 1,000
teacher handbooks 30 @ $2.50		75
video tapes for microteaching 30 1/2 hour tapes		450
Equipment purchase		
VTR equipment, 3 sets		6,000
	Subtotal	$7,525
Training costs		
Coordinator of training,		
6 weeks for planning and monitoring		2,025
**Substitutes to release teachers,		
1 substitute per 10 teachers × 5 weeks		
of training = 30 days @ $35/day		1,050
	Subtotal	3,075
	TOTAL	$10,600

* Based upon minicourse training costs.
** Optional; other ways may be used to release inservice teachers for microteaching; not required in preservice.

ing. Start-up costs such as purchase of training materials and video equipment are included as well as the cost of conducting the training itself. Once the training materials and equipment are available, this type of training runs approximately $100 per teacher. Given a shorter training period, a larger number of teachers to work with the coordinator, or less microteaching, the costs would be less. Ultimately, the cost would depend upon whether the training was part of a larger performance-based assessment and training system or a single, one-shot, effort.

Summary

In this paper questions have been raised regarding

—the purposes of assessment;
—the teacher performance to be assessed; and
—the individuals who will do the assessment, particularly the level of training and skill they must have based upon the assessment approach being used.

In each question area, several optional courses of action were presented. Different cost levels were associated with each option. Thus, the task for the individual, or agency, responsible for assessing teacher performance is to determine which options within each decision area best meet the policy-making and training information needs of that particular program (agency). Central to this decision will be the quantity and quality of data required given the purposes for which the findings will be used. A reasonable balance must be achieved between the desire to obtain detailed, highly reliable data and the total dollars to be spent.

Bibliography

Davis, O. L. and Tinsley, D. C. Cognitive objectives revealed by classroom questions asked by social studies teachers. Peabody Journal of Education, 1967, 45, 21-26.

Dodl, Norman, et al. "A Catalog of Teacher Competencies—A Working Document." Tallahassee: Florida State University, 1972.

Elementary Teacher Education Models, A Summary. Journal of Research and Development in Education, Vol. 2, No. 3. University of Georgia, Athens, Georgia, Spring 1969.

Gallagher, J. J. Expressive thought by gifted children in the classroom. Elementary English, 1965, 42, 559-568.

Guszak, F. J. Teacher questioning and reading. The Reading Teacher, 1967, 21, 227-234.

Herbert, John. A research base for accreditation of teacher preparation programs. Accreditation and Research Problems, J. L. Burdin and M. T. Reagan, eds. ERIC Clearinghouse on Teacher Education, Washington, D.C., 1971.

Houston, W. Robert. "Strategies and Resources for Developing a Competency-Based Teacher Education Program." University of Houston, New York State Education Department and Multi-State Consortium on Performance-Based Teacher Education, 1972.

Husek, T. R. and Sirotnik, Ken. Matrix sampling in educational research: an empirical investigation. Paper presented at the 1968 convention of the American Educational Research Association, Chicago, Illinois.

Joyce, Bruce R. Estimating costs of a competency orientation. Paper presented at Multi-State Consortium on Performance-Based Teacher Education, New Orleans, February 1973.

Joyce, Bruce R. and Weil, Marsha. "Models of Teaching." Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Lord, F. M. and Novick, M. R. "Statistical Theories of Mental Test Scores." Reading, Mass.: Addison-Wesley, 1968.

Rosenshine, B. "Teaching Behaviours and Student Achievement." London: National Foundation for Educational Research in England and Wales, 1971.

Silberman, C. E. "Crisis in the Classroom." New York: Random House, Inc., 1970.

Smith, B. O. Certification of educational personnel. University of South Florida, unpublished paper, 1971.

Turner, R. L. Relationships between Teachers for the Real World and the elementary models programmatic themes and mechanisms payoffs, mechanism and cost. In B. Rosner (ed.), "The Power of Competency-Based Teacher Education." Boston: Allyn and Bacon, Inc., 1972.

The Teachers Viewpoint

By

SANDRA FELDMAN

I have been asked to address my comments specifically to the concerns of State Education Agency personnel who have been and will be making significant policy decisions concerning performance education. I intend to do just that—and I hope you'll forgive me.

Although it doesn't seem to be, it should be a source of worry to state education departments that teachers are suspicious and wary of them, that they are not looked upon by the profession as allies, that they are not looked to as a resource for the solving of educational problems, that they are not thought of as providing educational leadership.

It should be food for thought, at least, that often the policymakers in state education departments, in the business of prescribing for the schools, find themselves on opposite sides of the barricades from teachers in legislative and other controversies on school matters, ranging from certification to teacher education to "alternative" school plans or performance contracting, or vouchers, or teacher evaluation, or professional practices acts, or questions of school structure and governance—or a host of other issues.

Why?

Why don't state education departments have a healthier respect for classroom teachers, for their accumulated experience and expertise, for their organizations as an expression of teacher concerns?

Why aren't teachers more significantly involved in the process of decision-making at state education department levels. (And by teachers I do not mean the state education department personnel who once were in the classroom and feel themselves able, therefore, to speak for teachers. Nor do I mean selected individual teachers, handpicked by school administrations. I mean teacher leaders selected democratically by the teachers themselves, through their organization.)

I would like to hear what you think the reasons are: I think they are manifold and I will tell you what I think first.

One reason, in my opinion, is an incomprehensible inability of education administration to accept the fact that times have changed, that teachers are organized, that they insist on speaking for themselves and on having a voice in the decisions which affect not just their working lives, but the education of the children they teach.

Insofar as your invitation to this conference of teacher representatives demonstrates some consciousness of that, I congratulate you—and I intend to take full advantage of this opportunity. But I do so with a certain skepticism; as a participant in many conferences such as this one, I seldom see the fruits of such exchanges in concrete application back home.

Another reason for a lack of real cooperation between teachers and their state education departments is a result, I think, of some state education department officials seeing themselves as in the vanguard of educational change and seeing teachers as defenders of the status quo. To teachers, this characterization is ludicrous. They do not see their demands for drastically reduced class size, for massively increased social, psychological, and health services for children, for the maintenance of high professional standards as "old hat." They do not see performance contracting, "accountability," legislation like the Stull Act, differentiated staffing, and merit pay schemes as revolutionary improvements, but as backward steps and attacks on the integrity of the profession. They are not impressed by paper-weighty master plans or glossy brochures advertising "bold new steps" which usually do nothing whatsoever to help them in the classroom.

There is yet another reason, in my opinion, for the "credibility gap" between teachers and state education departments; and this one I will give a little more time to, for it will lead me to the subject at hand. That is, I fear that many state education department officials in policymaking roles (1) are under great pressure to respond to the now famous "crisis of confidence" in the schools, without the knowledge and resources to do something real to end failure, and (2) share in the growing public attitude that teachers are responsible

for school failure and either don't know how to do something about it—or don't want to.

So, if we are to succeed in achieving the very worthy goal of this consortium—which is supposed to be "devoted to improving the quality of State Education personnel and their decision-making abilities," we have to first establish that one important, essential way of improving that policy decision making is to listen to the teacher organizations. Then we have to talk about the crisis in confidence that spurred the performance-based movement—why we have it and what we ought to do about it.

Sandra Feldman

Crisis in Confidence

The schools do not appear to be meeting the demands of a changed economy, of a job market which requires more and more skills and increasingly higher levels of training. Whether or not the schools in the past were the vehicle for upward mobility, for providing the path out of poverty, most people believe that they were—and expect them to continue that function. This problem of higher expectations is compounded by the social turbulence of the sixties and by the racial conflict which so often was most visible around school issues.

Second, despite the recent "revelation" by Christopher Jencks that schooling cannot end economic inequality among the adult population—and despite the fact that this was undoubtedly always true, that economic equality would have to be legislated as a matter of social policy—public schooling has undeniably produced a literate citizenry. Where in the past—and especially through the thirties, forties, and fifties—school teachers were among the most educated people in our society, even when they only had a high school or training school degree, today, the country is full of educated people. Thousands upon thousands of parents and citizens in white collar government employment, in advertising, in media jobs, accounting, computers, in a wide variety of nonteaching academic pursuits, feel that if they were not working at their present occupation, they could teach. Today, although teachers are still among the most educated, they do not have the status that comes with having knowledge that others do not have. While a holder of a Ph.D. in economics feels that a doctor's or a lawyer's skills are very special, even somewhat mysterious and intimidating, he does not feel that way about a teacher—even if the teacher also holds a Ph.D.

As a profession, education must, like other professions, assemble its experience and skill into a concrete, coherent body of knowledge as tangible as what exists in law or medicine.

That is why we must have performance-based teacher education. Not only will it make teacher education more relevant to classroom needs and thus improve education, but it will, if developed properly, provide us with that concrete knowledge without which we cannot much longer defend the public school system.

And that is why I believe, unshakeably, that developing a body of knowledge about the teaching-learning process ought to be the initial thrust and prime concern at present of the performance-based teacher education movement. We should be setting about doing what Fred MacDonald is proposing be done by a National Commission on Research and Development. We should be making a coordinated, long-term commitment to validating teacher competencies —not what is being done, which is a short-term commitment to listing them. We should be working at proving what teacher behavior, what teaching strategies, effect what learning and how.

We should not, as the New York State Education Department is doing, be telling teacher education institutions to "Poof! Change over to a performance-based teacher education program for certification approval purposes." At a meeting several months ago, to which representatives of teacher education institutions and the state teacher organization were invited, there were as

many different notions of what performance-based teacher education is as there were deans of education at the meeting. And the leadership in the state education department is not providing a model or a research design or any real guidance—unless you call requirements listed on a form for program approval, guidance. We will have as many different, and as many irrelevant and inadequate teacher education programs as we do right now—but New York State will "brag" about its performance-based teacher education program at conferences like these throughout the country.

Just recently, at a meeting of sophisticated chairmen of education departments of one of our largest teacher education institutions, I saw education department representatives of a college struggle with a paper description of what they hoped would be considered a performance-based program; and they had listed not teacher competencies, but behavioral objectives—which is fine to do, but from which, as Bunny Smith has emphasized over and over, you learn nothing about the teaching process.

I have grave doubts as to whether the states can autonomously develop meaningful performance-based programs—but I have no doubt that most teacher education institutions, without resources for research and without access to a large variety of classrooms, cannot do it.

If the performance-based teacher education movement is going to be meaningful, it cannot be handled as a public relations gesture by state education departments and tossed back to the colleges which have been floundering all along. Nor should the effort to develop a knowledge base of teacher competencies be tied to certification at this time, and certainly not to continuing certification schemes, because teachers will not stand for the destruction of tenure and job rights through renewable certification—especially when a review of performance is based on unvalidated lists of competencies.

If performance-based teacher education is going to be meaningful, it cannot be—as every education "innovation" usually is—doomed to success. In any other profession—take medicine—strategies are tried and tested and discarded if they don't work and the knowledge gained from systematic research into those strategies, whether they cure the disease or not, provides a continuous buildup of intelligence in the field. In education, we scurry to hide—or ignore—failure. That the research we've done so far is inconclusive, because sophisticated instruments to successfully interpret and measure teacher behavior are still in the early stages of development—and I know you've heard MacDonald, Rosenshine, and Schalock—should not be a source of embarrassment to us. It should provide us with the resolve to push that important work forward.

Instead, each state goes its own way, without a pooling and sharing—not just of knowledge, but of political power to get a commitment for funds to do the research we all know needs to be done. In New York State we have a very definite timetable. By—is it 1984?—we will have performance-based teacher education, so-called, and performance-based teacher certification throughout the State. But will we know any more about education or about teaching? I think not—unless the National Commission for Research and Development is ready to share their work with us—and they haven't got nearly enough money to do the work that has to be done. But that is the direction we should be taking.

What we should be doing—and actually the multistate consortium might well be an excellent vehicle for this—is coordinating a single effort, a national effort to develop a knowledge base. We should be collating the limited knowledge we now have, and we should be building a small number of teacher models based on available research and the opinion of experienced classroom educators. We should be developing in a broad cross-section of schools throughout the Nation a systematic assessment and data-gathering machinery which would enable us to compare and study teaching behavior and its effects on learning response in a prescribed variety of school environments. We should teach prospective teachers in the models created—not hundreds of different ones, but a few—so that we can control for effects, and we should have a research design built in so we can validate the competencies we have agreed on by studying those new teachers, on the job, where their education should continue during an internship-probationary period.

While I believe that the main thrust of performance-based teacher education should, in the beginning, be preservice, there must be involvement of experienced classroom teachers because they have a great deal of expertise to offer and because they will learn a great deal in the process of participating.

This past summer, The City University of New York came to UFT to seek cooperation in the development of competencies for teacher-training models. In a short time, we recruited over 300 interested teachers (who were paid a fee for their participation, as they should be). They provided a valuable contribution to the work, and they learned a great deal:

"I was forced to think out in a concrete role my role and skills as a teacher."

"I was forced to analyze myself and think of what really makes a good teacher."

"It is high time an emphasis was put on specific competencies rather than abstract theories."

"I came away more optimistic about the future of education."

(And school critics are always seeking a way to better teacher "attitudes"!)

State education departments should be joining with the teacher organizations to demand that National Institute of Education and its funds be put at our disposal; that it be used as a resource for the profession, not in the form of hundreds of diverse grants awarded in a scattered way throughout the country, but in a massive, concentrated effort to develop the profession in the way I have just briefly described.

As state education department personnel, you should not be plunging forward blindly, as I think you are, because of the great public pressure to "do something." You should insist on your right and your need to base your decision making on substantive, proven, professional knowledge. Whether you do or don't, the teacher organizations will insist that you do; and if there is a struggle between those demanding quickie public relations solutions and the teachers who are demanding substance, you will be squeezed in the middle —unless you have already chosen sides—and the promise of performance-based teacher education may die in the process.

Some Substantive Issues And Political Considerations In Performance-Based Teacher Education

By

BERNARD H. MCKENNA*

What Have We Heard?

We've heard here over the past 4 days that

1. the state of the art is little advanced—in fact someone termed it "primitive";
2. there is little data on what makes for competence—as some put it—"We know too little about what's worth teaching";
3. the most difficult job will be defining competencies;
4. adequacy of assessment apparatus is the most critical issue and will be most difficult to develop;
5. the definitions of behaviors are, at this point, almost totally hypothetical;
6. we lack summative evaluation tools; i.e., there are rating forms for making global judgments, and there are research tools containing specific skill formulation—but nothing in between;
7. there is a complex interaction between the setting and performance, an interaction that affects performance in a major way, one that we haven't much taken into account, let alone analyzed;
8. the affective domain is being neglected in performance-based teacher education;
9. the tools for measurement in the social sciences, including teacher education, are in an infant stage; i.e., even with the sophistication developed in measurement on cognitive skills such as reading, we're still not able to diagnose causes of all reading difficulties, let alone prescribe behaviors that will correct them—so, it has been asked, how will we soon be able to measure attitudes, values, and the like?
10. there is a monumental problem in matching teaching styles to learning styles;
11. finally we've heard that measuring specific performances at various criterion levels will be costly.

There are a lot of other lesser problems we've heard about, and likely some additional major ones I've omitted.

What Does What We've Heard Say To Us?

One thing it might say is "forget it." My starting off with such a laundry list could cause you to believe that's where I am—"forget it," or that I'm launching into a defensive diatribe.

Actually, I believe that performance-based teacher education is a promising and viable concept and should be pursued.

My listing these concerns is for the purpose of putting a framework around what I'm about to respond to, and that is the following question.

What Are the Main Issues in Performance-Based Teacher Education That Researchers, Developers, Teachers, Professors, and Others Should Be Emphasizing?

In my judgment, a first line of activity should be around tools for evaluation. Performance-based teacher education will rise or fall on our ability to evaluate. That is, a first-order question is 'How shall we determine that teacher candidate A is now ready to practice the profession and should be licensed to do so and that candidate X should "go back to the drawing board"?'

Another way of saying it is that until we can measure performance in such a way that we are able to confidently identify minimal levels of performance to practice the profession, we will not have performance-based teacher education.

Some have said at this conference that prior to identifying minimal levels of performance there are the more difficult jobs of

* Dr. McKenna is professional associate, Instruction and Professional Development, National Education Association.

(a) establishing goals and objectives for education
(b) identifying performances (competencies) required for achieving the goals and objectives.

I don't believe those to be the most difficult tasks; and even if they were, we're further along in accomplishing them than we are on the evaluation matter.

For example, there are available

1. some well-developed needs-assessment systems;
2. several quite complete taxonomies;
3. the works of Mager, Atkin, Popham, and others on getting goals and objectives into more nearly manageable performance-based terms;
4. "The Florida Catalog of Competences" and "Signs of Good Teaching" (the latter is less known but highly developed);
5. the Seven Cardinal Principles or Ten Purposes of Secondary Education;
6. the techniques of most recent vintage reported at this conference by Adele Thomas of The City University of New York in which they were able to gain high consensus on which skills should be most valued;
7. the Northwest Regional Laboratory materials and those of the Teacher Corps.

I'm not that concerned about our ability to arrive at goals and objectives.

Some in this conference have argued that "instructional systems," as Bruce Joyce termed them, will be most difficult to develop. I'd argue that we have made quite satisfactory progress on these also.

Whether one goes with those toward the affective end of the continuum like "Man a Course of Study" and the Taba Social Studies Curriculum or the more behavior-modification oriented types reflected in some of the packaged curriculums, there is a variety of quite well-developed instructional systems to choose from which include both content and teaching strategies.

So that brings us back to evaluation devices—and I repeat—performance-based teacher education will rise or fall on our ability to evaluate performance in relation to agreed-upon standards.

Past research on evaluation leaves much to be desired; and I'm convinced that too little emphasis is being given to research and development in this key area in present efforts with performance-based teacher education.

There are places to start, and I'm surprised that more of the programs haven't attended to those promising possibilities. For example, the performances identified by Rosenshine and Furst, such as clarity, variability, enthusiasm, have shown some promising relationships to outcomes with students and are one place to begin.

There are also others that should be considered whether or not they show direct relationship to learning outcomes. The point here is that there are some processes which may not lead directly to agreed-upon goals (or at least cannot be demonstrated to do so) which should be promoted and which are worthy of evaluation for their own sake. One brief illustration of this. A recent study of several thousand students representative of the American high school population concluded that great numbers of students may be developing little affinity to the democratic process simply because they have little opportunity to experience it during their school career. If this is so, is it not important that the process of education in the schools become, in as much as possible, a microcosm of the best of democracy as it is practiced in the greater society? And isn't this important whether or not such processes can be definitively shown to contribute to specific learning outcomes? Evaluation systems for performance-based teacher education should incorporate such criteria.

I believe, as Bruce Joyce predicted earlier in this conference, that it will take 20 or more years to develop reliable and valid evaluation tools. It will probably cost more than his suggested $100 million. When asked if it were possible to develop an accountability model for New York City, Henry Dyer of Educational Testing Service, responded that it could be done but would be more difficult than getting to the moon—and it is well known that NASA had for several years for that project, almost unlimited resources and unlimited power.

Three years later Fred McDonald and his people have made some progress on the New York City accountability mandate. I've just read the description of their model. It appears promising on first look. But it almost totally skirts the problem of relating performance to student achievement. What might be the cost and the time line if this issue were responded to?

Nat Gage told us on Sunday that Lindquist gave warning 30 years or more ago that there are some 18 or 20 variables other than teacher performance which contribute to student achievement. Gene Glass put it well recently when he concluded from his research that "aside from the irrelevance of much of the content of standardized achievement tests, their use in evaluating teachers is unjust."

If you don't think that kind of thing is being promoted, you have only to read the Stull Act in California. It's like directing the medical profession to find a cancer cure in a year or mandating the Nation's economists to correct the balance-of-payments conundrum in a few months. Or look at the Fleischmann Commission Report in New York State, or some of the Management by Objectives programs that are being laid on school staffs across the country.

Bernard H. McKenna

Two other quick things on the evaluation issue. The first is that where work is being done on evaluation related to performance-based teacher education, there appears to be overemphasis on evaluation of didactic teaching styles. One hears the terms "teaching lessons" and "collecting student papers" and other lecture-type activities referred to all too frequently. I thought Betty Ward was very bold when she raised the question yesterday about the desirability of assessing skills of total groups since the teaching process is becoming more of a team arrangement, more interpersonal, more a reflection of the open-classroom concept, more guided independent study. I say she was bold, since the minicourses she has been so involved in developing to a considerable degree assume a didactic teaching style.

The second point is that there should be more consideration of evaluating staff as a team. I think it was Betty Ward also who suggested "don't test every skill in every teacher." In this regard, we might look on the staff as an orchestra—evaluating for complementary skills—not sameness. (An orchestra of all piccolo players would be pretty dull and would certainly receive a strange review from the music critics.) There has been considerable research done on this team concept of evaluation over a period of 30 years in the Institute of Administrative Research, Columbia University. "Indicators of Quality" is a promising tool for this purpose.

How Does Performance-Based Teacher Education Relate to Inservice?

If inservice education had ever amounted to anything, we'd have had respectable performance-based teacher education long ago.

We've had inservice education a long time. We've also had performance evaluation inservice a long time. In fact, one might expect most inservice evaluation to be performance-based and for the purpose of determining corrective measures (inservice needs).

Instead this is the way it has been: inservice evaluation of school staffs has been poorly conceived, little developed, only haphazardly implemented and the results often inappropriately used.

In our work with 1-million plus teachers across the country, we note some quite commonly held attitudes on their part about inservice education:

1. They find it threatening.
2. They assume its main purpose will be to determine change of status—retention, tenure, promotion.
3. They feel uninformed about the criteria by which they are evaluated.
4. They assume evaluation will be accomplished through brief observations (too brief) and based on some kind of checklist.
5. They expect it will be done to them by somebody else, someone in a more elevated position.
6. They say the feedback they get from it is inadequate and often absent.
7. They believe there will be few concrete recommendations for improvement as a result of their being evaluated.
8. They expect that following evaluation there will be little or no help to improve their performance through inservice programs geared specifically to correct the lacks identified through evaluation.

That's the way it is according to our members. So if performance-based teacher education is to be carried over to inservice education, it will need to respond to all the deficiencies of present inservice evaluation and inservice education. And that's a big order.

If performance-based teacher education works, with its proposed field orientation, hospital-schools concept,

internships, and career ladders, then the transition from preservice to inservice both in terms of the process and the evaluation of it should become so gradual, so smooth, that one can't tell where one leaves off and the other starts. It would then approach what the law profession is getting underway in the new Antioch Law School in the city where I live, and what Western Reserve University got underway in medical education more than 15 years ago.

That's short shrift for the inservice aspects, but there is some other ground to cover.

What Is Recommendable?

A. Involve on a parity-basis—not just token—all groups within the profession who will be affected by the program. You might begin with some of the states in this consortium.
B. Assure that goals and objectives for PBTE are based on goals and objectives for schooling.
C. Assure that goals and objectives for schooling are cooperatively developed by parents, educators, and students.
D. When objectives are turned into performance criteria, provide for macro as well as micro performances.
E. Assure that program and conditions are evaluated concurrently with performance and that they are related to it. It's futile to evaluate a teacher's performance without also evaluating all these conditions that make competent performance possible.
F. Involve all those who will be affected by it, in developing the evaluation plan—prospective teachers, teacher supervisors, their students, and some parents.
G. Use multiple indexes for evaluation.
H. Involve a variety of personnel in conducting evaluation—self, peers, supervisors, students.
I. Thoroughly train those who will evaluate. This may be the most important.
J. Develop concurrently plans for correcting both those program and condition elements and performances which are shown to be deficient as a result of evaluation. This may be the most costly, particularly on the inservice side.
K. Assure that there will be ample opportunity to improve (including direct assistance, time, and materials) for those judged to require improvement as a result of evaluation.
L. Provide for full-blown, written, agreed-upon arrangements for both substantive and procedural due process when evaluation systems are to be used for decision making on other than determining inservice; that is, the retention, tenure, promotion.
M. Secure research funds commensurate with the largeness of the task.
N. Develop time lines that allow for realistic accomplishment of what's accomplishable, taking into account the state of the art and the complex interrelationships of the many variables as well as the feebleness of tools in social science measurement.
O. Researchers get in there and testify before legislatures on what's possible and what's not.
P. Stay close to the schools. (I couldn't disagree more with Fred McDonald when he said, "You can't learn anything in the messy classroom situation." That may be why ETS was so long in recognizing the injustices of the National Teachers Examination.)
Q. And finally, provide sufficient periods for test and tryout, evaluation, and recycling before dissemination—before making extravagant promises about all the ills the new program is sure to correct. Let us not allow this innovation to suffer the same setbacks as:
   1. the teacher aide concept which got underway in Bay City, Michigan, 20 years ago, but was probably delayed 10 or more years by poor project design and overstated promotion;
   2. or educational television which suffered some of the same problems partially as a result of the Hagerstown experiment in the midfifties;
   3. or differentiated staffing which appears to have suffered in a similar way in the last 2 or 3 years.

What Are The Politics Of The Matter?

My final charge was to deal with the politics of performance-based teacher education.

As you have doubtless noted by now, I assumed I had several charges.

Ted Andrews told me by phone that the symposia were to be on assessment in CBTE. Later in a letter he suggested I deal with its relation to inservice education. And when the program came out, the term "politics" appeared in the title of my presentation.

According to someone's definition, politics is "the art of the possible." If that's so, I've already covered what I think is possible and what conditions I believe are necessary to accomplish the possible. Just an additional thought on political considerations.

I am weary of power struggles. Unlike Pat Goralski, I don't accept them as givens. I hope soon we'll get beyond them so we can concentrate our energies on getting the priorities of the American people in better balance in order to obtain the conditions and resources that are so badly needed for schooling, all the way from prenursery school to graduate programs.

This can be accomplished, at least partially, if we work toward parity in decision making and then concentrate on issues rather than on which individuals or groups will have the most power.

Redistribution of power so that it will be more equitable among all groups concerned will in the end give us all more power.